CHALMERS

Space-Time Parameter Estimation in Radar Array Processing

FREDRIK ATHLEY

Department of Signals and Systems
School of Electrical Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2003

Thesis for the degree of Doctor of Philosophy

Technical Report No. 452

Space-Time Parameter Estimation in Radar Array Processing

by

Fredrik Athley

CHALMERS

Department of Signals and Systems
School of Electrical Engineering
Chalmers University of Technology
Göteborg, Sweden 2003

Space-Time Parameter Estimation
in Radar Array Processing
Fredrik Athley
ISBN 91-7291-300-2

This thesis has been prepared using LaTeX.

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 1982
ISSN 0346-718X

Technical Report No. 452
School of Electrical Engineering
Chalmers University of Technology
ISSN 1651-498X

Signal Processing Group
Department of Signals and Systems
School of Electrical Engineering
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
Telephone: + 46 (0)31-772 1000

Front cover: The surface is the asymptotic maximum likelihood criterion function for the problem of finding the directions of arrival with a sensor array. The highest peak in this function is at the true directions. The other peaks are sidelobes due to near-ambiguities in the spatial transfer function. For a finite number of measured data samples, noise may cause one of these sidelobes to become the global maximum, resulting in a large estimation error. The estimates obtained by finding the global maximum in a noise-corrupted version of this function are plotted as red dots. Most estimates are around the mainlobe but there are also clusters around the sidelobe peaks. See Chapter 10 in this thesis for further details.

Printed by Vasastadens Bokbinderi
Göteborg, Sweden, May 2003

To Johanna, Jonathan, and Emelie

Abstract

This thesis is about estimating parameters using an array of spatially distributed sensors. The material is presented in the context of radar array processing, but the analysis could be of interest in a wide range of applications such as communications, sonar, radio astronomy, seismology, and medical diagnosis. The main theme of the thesis is to analyze the fundamental limitations on estimation performance in sensor array signal processing. To this end, lower bounds on the estimation accuracy as well as the performance of the maximum likelihood (ML) and weighted least-squares (WLS) estimators are studied.

The focus in the first part of the thesis is on asymptotic analyses. It deals with the problem of estimating the directions of arrival (DOAs) and Doppler frequencies with a sensor array. This problem can also be viewed as a two-dimensional (2-D) frequency estimation problem. The ML and WLS estimators for this problem amount to multidimensional, highly nonlinear optimization problems which would be expensive to solve in real-time in a radar system. Therefore, simplifications of this problem are of great interest. It is shown in this thesis that, under some circumstances, the 2-D problem decouples into 1-D problems. This means a dramatic reduction in computational complexity with insignificant loss of accuracy.

The second part contains a performance analysis of the ML DOA estimator under conditions of low signal-to-noise ratio (SNR) and a small number of data samples. It is well known that the ML estimator exhibits a threshold effect, i.e. a rapid deterioration of estimation accuracy below a certain SNR. This effect is caused by outliers and is not captured by standard analysis tools. In this thesis, approximations to the mean square estimation error and probability of outlier are derived that can be used to predict the threshold region performance of the ML estimator with high accuracy. Moreover, these approximations alleviate the need for time-consuming computer simulations when evaluating the ML performance.

Keywords: Sensor arrays, direction of arrival, Doppler frequency, two-dimensional frequency estimation, radar, lower bounds, sparse arrays, threshold effect, maximum likelihood, weighted least squares.

Contents

Acknowledgments

This part of the thesis is probably the one that is read by most people. Unfortunately, since procrastination has been a salient feature of my work as a PhD student and this section is the last thing to write, there is not much time left before I need to run to the printing office. Nevertheless, I would like to seize this opportunity to thank some of the people who have meant a lot to me during my graduate studies; financially, intellectually, or emotionally.

My work has been funded by Ericsson Microwave Systems under the *Intelligent Antenna Systems* research programme. I would like to thank Ericsson for so generously sponsoring my graduate studies. Special thanks to Dr. Ragnar Arvidsson and Ingmar Karlsson who enrolled me in the research programme. I would also like to thank Prof. Anders Derneryd for coordinating the project and Dr. Christer Engdahl for fruitful discussions related to Part II of this thesis.

Next, my thanks go to my supervisor Prof. Mats Viberg for his encouragement and for sharing some of his brilliant expertise in signal processing. When you sometimes get stuck in problems that seem more or less hopeless, some people have the ability to lift you to a higher level and make everything seem so easy. Mats is certainly one of them. He has also the important ability of putting together a research group consisting of people that, in addition to being excellent researchers, you can have great fun with.

I would like to thank all the people in the Signal Processing, Image Analysis, Circuit Design, and Medical Electronics groups for enjoyable coffee/tea breaks, lunches, and various after-work activities. I will really miss the nice and relaxed atmosphere we have in our corridor. Special thanks to Prof. Mats Viberg, Dr. Chris Brown, Patrik Bohlin, and Mikael Tapio for proofreading the thesis. Having my manuscript pass this eye of the needle feels safe for the defense.

Finally, and most importantly, I would like to thank my family for love and inspiration. In particular, I thank my wife, Johanna, for an immense

amount of support and patience in the course of writing this thesis, and our children, Jonathan and Emelie, for just simply being the most important part of my life. I hope that all the effort behind this thesis has not just been for my own good, but that it will give, in one way or the other, something in return to you.

Abbreviations and Acronyms

1-D	One-Dimensional
2-D	Two-Dimensional
AP	Alternating Projection
BB	Barankin Bound
CPI	Coherent Processing Interval
CRB	Cramér-Rao Bound
DFT	Discrete Fourier Transform
DML	Deterministic Maximum Likelihood
DOA	Direction Of Arrival
EXIP	Extended Invariance Principle
FIM	Fisher Information Matrix
FFT	Fast Fourier Transform
LS	Least Squares
MAP	Maximum A Posteriori
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimator
MMSE	Minimum Mean Square Error
MRA	Minimum Redundancy Array
MSE	Mean Square Error
MTI	Moving Target Indication
PDF	Probability Density Function
PRI	Pulse Repetition Interval
Radar	Radio Detection and Ranging
RCS	Radar Cross Section
RMS	Root Mean Square
RMSE	Root Mean Square Error
SML	Stochastic Maximum Likelihood
SNR	Signal to Noise Ratio

STAP	Space-Time Adaptive Processing
SVD	Singular Value Decomposition
ULA	Uniform Linear Array
WLS	Weighted Least Squares
WWB	Weiss-Weinstein Bound

Notation

In the thesis, matrices and vectors are set in boldface, with upper-case letters used for matrices and lower-case letters for vectors. The meaning of the following symbols are, if nothing else is explicitly stated:

$\mathbf{A}^T$	Transpose operator.
$\mathbf{A}^*$	Complex conjugate.
$\mathbf{A}^H$	Hermitian transpose, i.e. complex conjugate transpose.
$\mathbf{A}^\dagger$	The Moore-Penrose pseudo-inverse of $\mathbf{A}$, i.e. $\mathbf{A}^\dagger = \left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H$.
$\mathbf{P}_{\mathbf{A}}^{\perp}$ or $\mathbf{\Pi}_{\mathbf{A}}^{\perp}$	Orthogonal projection matrix onto the null space of $\mathbf{A}^H$, i.e. $\mathbf{P}_{\mathbf{A}}^{\perp} = \mathbf{\Pi}_{\mathbf{A}}^{\perp} = \mathbf{I} - \mathbf{A}\mathbf{A}^\dagger$.
$\mathbf{A}^{1/2}$	Hermitian (i.e. $(\mathbf{A}^{1/2})^H = \mathbf{A}^{1/2}$) square-root factor, i.e. $\mathbf{A} = \mathbf{A}^{1/2}\mathbf{A}^{1/2}$.
$\mathrm{Tr}\{\mathbf{A}\}$	Trace operator.
$\lvert\mathbf{A}\rvert$	Matrix determinant.
$\mathrm{vec}(\mathbf{A})$	The column vector that is obtained by stacking the columns of $\mathbf{A}$.
$\{\mathbf{A}\}_{ij}$ or a_{ij}	The (i,j)-th element of the matrix $\mathbf{A}$.
$\lVert\mathbf{A}\rVert_F$	The Frobenius matrix norm, defined as $\lVert\mathbf{A}\rVert_F^2 = \mathrm{Tr}\{\mathbf{A}\mathbf{A}^H\}$.

$\mathbf{A} \otimes \mathbf{B}$ — The Kronecker product. For an $N \times M$ matrix $\mathbf{A}$ and a $K \times L$ matrix $\mathbf{B}$ it is defined as the $NK \times ML$ matrix

$$\mathbf{A} \otimes \mathbf{B} \triangleq \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1M}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{N1}\mathbf{B} & \cdots & a_{NM}\mathbf{B} \end{bmatrix}$$

$\mathbf{A} \diamond \mathbf{B}$ — The Khatri-Rao product, or column-wise Kronecker product. For an $N \times M$ matrix $\mathbf{A}$ and a $K \times M$ matrix $\mathbf{B}$ it is defined as the $NK \times M$ matrix $\mathbf{A} \diamond \mathbf{B} \triangleq \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \ldots & \mathbf{a}_N \otimes \mathbf{b}_N \end{bmatrix}$ where $\mathbf{a}_k$ denotes the k-th column of the matrix $\mathbf{A}$.

$\mathbf{A} \odot \mathbf{B}$ — The Hadamard or Schur product, i.e. element-wise multiplication.

$\mathbf{A} > \mathbf{B}$ — The matrix $\mathbf{A} - \mathbf{B}$ is positive definite. Furthermore, $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite.

$\mathrm{diag}\,(a_1, \ldots, a_K)$ — A $K \times K$ diagonal matrix with diagonal elements a_1 through a_K.

$\mathrm{Diag}\,\{\mathbf{A}\}$ — A column vector containing the diagonal elements of the matrix $\mathbf{A}$.

$\mathbf{I}_m$ — The $m \times m$ identity matrix. Frequently, the subscript m is omitted when there is no risk of confusion.

$\hat{\theta}$ — An estimate of the parameter θ.

θ_0 — The true value of the parameter θ.

$\mathrm{Re}[\cdot]$ — The real part of a complex quantity.

$\mathrm{Im}[\cdot]$ — The imaginary part of a complex quantity.

$\mathrm{E}[\cdot]$ — The expectation of a random variable.

$a_n = o(g_n)$ — a_n is of smaller order than g_n, i.e. $\lim_{n\to\infty} g_n^{-1} a_n = 0$.

$a_n = O(g_n)$ — a_n is at most of order g_n, i.e. there exists a real number M such that $g_n^{-1}|a_n| \leq M$ for all n.

$a_n = o_p(g_n)$ — The random variable a_n is of smaller order in probability than g_n, i.e. for every $\epsilon > 0$ $\lim_{n\to\infty} P(g_n^{-1}|a_n| > \epsilon) = 0$

$a_n = O_p(g_n)$	The random variable a_n is at most of order in probability g_n, i.e. for every $\epsilon > 0$ there exists a real number M such that $P(\|a_n\| \geq Mg_n) \leq \epsilon$ for all n.
$\mathbf{x} \in \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$	The random vector $\mathbf{x}$ is distributed as a normal random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$.
$\mathbf{x} \in \mathrm{As}\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$	The random vector $\mathbf{x}$ is asymptotically distributed as a normal random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$.
$\delta_{t,s}$	The Kronecker delta; $\delta_{t,s} = \begin{cases} 1 & t = s \\ 0 & t \neq s \end{cases}$
$\arg\min_x f(x)$	The minimizing argument of the function $f(x)$.
$\arg\max_x f(x)$	The maximizing argument of the function $f(x)$.
$\log(x)$	The natural logarithm of x.
$\bigcup_{n=1}^{N} A_n$	Union of the events $A_n, n = 1, \ldots, N$.
$\bigcap_{n=1}^{N} A_n$	Intersection of the events $A_n, n = 1, \ldots, N$.

Chapter 1

General Introduction

1.1 Background

This thesis is about estimating unknown parameters using an array of spatially distributed sensors. Sensor arrays are used in a wide range of applications such as radar, passive sensors, sonar, communications, seismology, radio astronomy, medical diagnosis and chemical analysis. Sensor array signal processing has received tremendous interest in the literature over the past few decades, see e.g. [Far92, Hay95, HLE93, HRKV92, JD93, KV96, MM80, Nic87, Tre02] and the references therein. The motivation behind the work presented in this thesis stems from radar applications, but the models and techniques treated could be applicable to a wide range of applications.

More specifically, this thesis deals with the problem of estimating the spatial and temporal frequencies of plane waves impinging on an array of antenna elements. This may also be viewed as a two-dimensional (2-D) frequency estimation problem. In radar applications the spatial frequency is related to the direction of arrival (DOA) from an object that is backscattering a fraction of the transmitted electromagnetic energy. The temporal frequency, or the *Doppler frequency*, is related to the relative speed between the object and the radar platform via the well known Doppler effect.

In this chapter, a brief introduction to radar systems is first given. A parameterized mathematical model for the output of a sensor array is then derived. Further, we discuss different signal and noise models that are pertinent to the problems considered in this thesis. We also give a concise overview of the basic estimation methods used in sensor array signal processing and lower bounds on the estimation accuracy. The outline and the contributions of the thesis are then summarized. Finally, some suggestions for future research are provided.

1.2 Pulsed Doppler Radar

In this section, a brief introduction to radar systems is given, with particular emphasis on the aspects that are important for reading this thesis. The interested reader is referred to [Bar88, NRC91, Sko86, Sko89] for a more comprehensive treatment.

Radar systems are used for detecting objects and estimating their positions by transmitting electromagnetic signals and analyzing the backscattered signals from the objects. Estimation of a object's range is enabled by transmitting short pulses and/or coded waveforms. A pulsed Doppler radar transmits a coherent burst of short pulses of a certain carrier frequency. The electromagnetic energy is concentrated to a particular spatial sector by transmitting with a focused beam. The beam can be created by a continuous aperture, e.g. a parabolic reflector, or an array of antenna elements.

An object in the radar's field of view backscatters a fraction of the transmitted energy, which is related to the object's *radar cross section* (RCS) and the distance between the radar and the object. If the object or the radar platform is in motion, there will be a *Doppler shift* of the carrier frequency, proportional to the radial velocity between the object and the radar. An object's DOA is traditionally determined by scanning a narrow beam over the search volume and monitoring the received signal energy. If the energy exceeds a certain threshold value, an object is detected and the DOA is given by the pointing direction of the beam. The distance to the object can be determined from the round trip delay of the transmitted pulse and the Doppler frequency may be estimated from the phase shift between successive pulses.

Beam scanning can be achieved either mechanically by a rotating reflector antenna or electronically by a phased array. A phased array consists of many small antenna elements with a phase shifter behind each element. By progressively shifting the phase from one array element to the next, a beam is formed in a direction determined by the size of the phase shift. The main advantage with electronic beam steering is that it provides high agility and flexibility.

Phased arrays have been used in radar systems since the 1950s. More recently, radar systems employing digital arrays have emerged. In a digital array, there is a receiver and an A/D converter behind each antenna element. Figure 1.1 depicts a signal processing view of the receiving part of a phased array and digital array system, respectively. For more details on phased arrays, the interested reader is referred to [Mai94].

In a digital array the beam steering can be performed by digital signal processors by simply multiplying the signal from each antenna element by a complex weight and then summing. Digital arrays provide even more

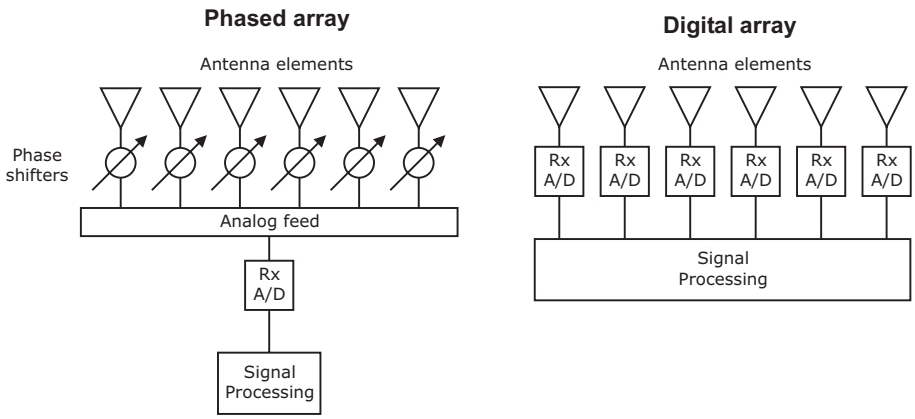

Figure 1.1: Phased and digital arrays.

flexibility than phased arrays since the complex weights need not produce simple, progressive phase shifts. For instance, directional interference can be rejected by computing weights such that the array beampattern has a null in the interference direction. More generally, a digital array provides the signal processors with spatial samples that can be used in more advanced parameter estimation methods.

In this thesis, radar systems employing multiple receive antenna elements are considered. It is assumed that several digital outputs are available from the antenna array, either from analog subarrays or from each antenna element. Thus, the receive antenna performs spatial sampling of impinging electromagnetic wavefronts. This enables digital signal processing in the spatial dimension.

The received signals are down-converted to baseband, matched filtered on a pulse-to-pulse basis and sampled by means of A/D converters. The samples thus obtained correspond to different distances to potential objects and are often referred to as *range bins*. This multidimensional data set can be visualized as a *data cube*, see Figure 1.2. Usually, we focus on a particular range bin and thus consider a 2-D space-time slice in the data cube, as illustrated in Figure 1.2. This slice consists of the spatial samples for all pulses at the current range bin. Within this 2-D subset we can perform spatial processing over the antenna elements and temporal processing over the pulses. The processing can be performed by joint 2-D methods or decoupled 1-D methods.

Once residing in the 2-D space-time slice, the first problem is to decide if one or several objects are present in the selected range bin. If a detection is declared, different signal parameter estimates are produced that carry information about the objects. Such parameters include DOA, range, RCS,

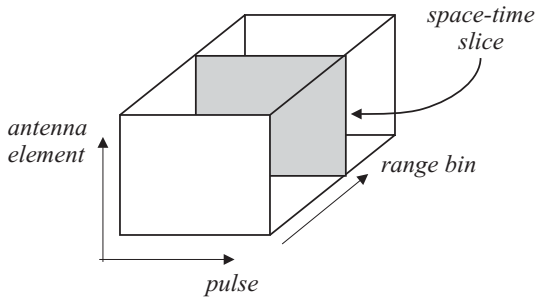

Figure 1.2: The radar data cube.

Doppler frequency, etc. Object parameter estimates are then fed to a Kalman filter bank used for tracking of the detected objects. In this thesis only the estimation problem is considered. More specifically we are interested in finding estimates of the DOAs and Doppler frequencies. Thus, we assume that the number of objects within a particular range bin has been estimated by some method, see e.g. [VOK91, WK85, Eri02].

The Doppler frequencies and DOAs are traditionally determined by performing temporal and spatial Fourier transforms respectively. In effect, a parabolic reflector antenna can be viewed as an analog spatial Fourier transform. The estimates are obtained from the peak in the corresponding periodograms. In radar terminology these processing stages are termed Doppler filter bank and beamformer respectively. However, it is well known that Fourier methods have a resolution that is limited to the inverse of the data length. For the problem at hand, this translates to a Doppler resolution limited to the filter width in the Doppler filter bank, and angular resolution limited to the antenna beamwidth.

The limited resolution of Fourier based estimation methods is essentially due to that they are designed to estimate the parameters of one object at a time. Therefore, these methods will have poor resolution in a multiple object scenario. By utilizing a properly parameterized mathematical model of the multiple object scenario, the resolution can be improved dramatically (at least in theory). Estimation methods based on these more refined models are often referred to as *model-based signal processing*[1].

Although model-based methods can also be applied to phased arrays and reflector antennas [DL97], a digital array is required to take full advantage of the potential of these methods. It is of great interest to study model-based signal processing methods that do not suffer from the resolution limitations

[1]One may think of the Fourier methods as model-based methods for single object scenarios.

of the traditional methods used in radar systems.

1.3 Sensor Arrays

In this section, a brief description of the data model used for the sensor array processing problem is given. A more detailed account can be found in ,e.g., [JD93, MM80, SM97, Tre02]. First we will present the basic spatial data model. This is a fairly general model that can be used in many applications. We then turn our attention to signal and noise models which are more application specific.

1.3.1 Spatial Data Model

Consider the scenario in Figure 1.3, in which K sensors receive waveforms from M point sources that are emitting, or backscattering, electromagnetic energy. The sources are assumed to be in the far field of the array, so that the

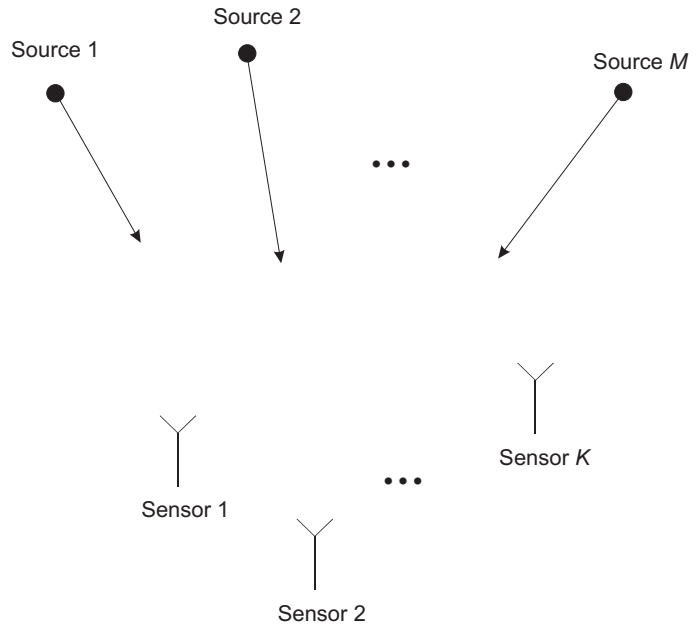

Figure 1.3: A sensor array receiving plane wavefronts.

wavefronts at the array can be considered to be planar. The sensors could be arranged arbitrarily in 3-dimensional space, but for ease of exposition it is assumed that the array and the sources are coplanar. This means that the directions to the sources can be characterized by the azimuthal angle

only. Furthermore, it is assumed that the signal waveforms are narrowband around a known carrier frequency.

We now derive a parameterized mathematical model of the sensor array outputs. To this end, consider an emitter signal $\bar{x}_r(t)$ and express it as a modulated carrier according to

$$\bar{x}_r(t) = \alpha(t)\cos[\omega_c t + \phi(t)], \tag{1.1}$$

where $\alpha(t)$ is the amplitude, ω_c the carrier frequency and $\phi(t)$ the phase. By narrowband we mean that $\alpha(t)$ and $\phi(t)$ vary slowly relative to the propagation time τ across the array, i.e.

$$\alpha(t-\tau) \approx \alpha(t) \quad \text{and} \quad \phi(t-\tau) \approx \phi(t). \tag{1.2}$$

The analytic signal representation of $\bar{x}_r(t)$ is given by

$$\bar{x}(t) = \alpha(t)e^{j\phi(t)}e^{j\omega_c t}. \tag{1.3}$$

The information about the direction to an emitter is contained in the time-delays of the propagating wave between the different sensors. The narrowband assumption implies that

$$\begin{aligned} \bar{x}(t-\tau) &= \alpha(t-\tau)e^{j\phi(t-\tau)}e^{j\omega_c(t-\tau)} \approx \alpha(t)e^{j\phi(t)}e^{-j\omega_c\tau}e^{j\omega_c t} \\ &= e^{-j\omega_c\tau}\bar{x}(t). \end{aligned} \tag{1.4}$$

This means that the narrowband assumption allows a time-delay of the signal to be modeled as a simple phase-shift of the carrier frequency. In each individual sensor, the received signal is demodulated by multiplying $\bar{x}(t)$ with $e^{-j\omega_c t}$. The signal

$$x(t) = \alpha(t)e^{j\phi(t)} \tag{1.5}$$

thus obtained, is referred to as the *complex baseband signal*.

Now, let $x(t)$ be the continuous-time complex baseband signal at some reference point. Without loss of generality, this reference point is assigned to the position of the first sensor. The noise-free output from the k-th sensor is modeled by

$$y_k(t) = h_k * x(t-\tau_k), \tag{1.6}$$

where h_k is the impulse response of the k-th sensor, $(*)$ denotes convolution and τ_k is the travel time of the wavefront from the reference point to the k-th sensor. This travel time is a function of the DOA, θ, of the wavefront, i.e.

$\tau_k = \tau_k(\theta)$. The narrowband assumption implies that the Fourier transform, H_k, of h_k is constant over the signal bandwidth. Under this assumption we get

$$y_k(t) = H_k(\theta)e^{-j\omega_c\tau_k(\theta)}x(t), \tag{1.7}$$

where the argument θ in $H_k(\theta)$ accounts for a possible variation of the amplitude and phase response with the DOA θ. The sensor outputs are collected in the complex $K \times 1$ vector

$$\mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_K(t) \end{bmatrix} = \mathbf{a}(\theta)x(t), \tag{1.8}$$

where

$$\mathbf{a}(\theta) = \begin{bmatrix} H_1(\theta)e^{-j\omega_c\tau_1(\theta)} \\ \vdots \\ H_K(\theta)e^{-j\omega_c\tau_K(\theta)} \end{bmatrix}. \tag{1.9}$$

The complex $K \times 1$ vector $\mathbf{a}(\theta)$ is referred to as the *steering vector*. It represents the array response to a unit wavefront from the DOA θ. The collection of vectors $\mathbf{a}(\theta)$ as θ varies over the parameter space of interest is termed the *array manifold*. The array manifold is assumed to be known. In practice, this means that the array has to be calibrated. The array manifold is said to be *unambiguous* if, for any collection of K distinct DOAs, the corresponding steering vectors are linearly independent. This property is important for guaranteeing unique estimates.

The sensors are assumed to be linear systems. Therefore, the above model for a single signal is readily extended to the case of M signals using superposition:

$$\mathbf{y}(t) = \sum_{m=1}^{M} \mathbf{a}(\theta_m)x_m(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(t), \tag{1.10}$$

where

$$\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{a}(\theta_1) & \dots & \mathbf{a}(\theta_M) \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1(t) \\ \vdots \\ x_M(t) \end{bmatrix}. \tag{1.11}$$

Due to noise, modeling errors, etc, the model in (1.10) can never explain the observed data exactly. To account for these effects, we include an additive noise term $\mathbf{n}(t)$ in the array output model:

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(t) + \mathbf{n}(t). \tag{1.12}$$

Via suitable receiver equipment and A/D converters, a set of N temporal samples of the K-vector $\mathbf{y}(t)$ is assumed to be available. Similar to the antenna characteristics, the effect of receiver filters etc. can be included in the steering vectors, as long as the model is still linear. The temporal samples are denoted (with some abuse of notation) $\mathbf{y}(n)$, and are taken at discrete time instants $t_n,\ n = 1, \ldots, N$. In the case of uniform sampling, $t_n = (n-1)T_s$, where T_s is the sampling time. The samples collected at a particular time instant in the vector $\mathbf{y}(n)$ are often referred to as a *snapshot*.

To illustrate the parameterization of the array manifold, consider a uniform linear array (ULA) with K identical sensors and uniform spacing, d. The DOA, θ, is defined as the angle relative the normal of the array axis. The normal of the array axis is also referred to as *boresight* or *broadside*. From Figure 1.4 it is readily seen that the propagation delay, τ_k, between the first and the k-th sensor is given by

$$\tau_k = \frac{(k-1)d\sin\theta}{c},\ k = 1, \ldots, K, \tag{1.13}$$

where c is the propagation speed. If we assume that the individual sensors are isotropic and have unity gain, i.e. $H_k(\theta) = 1, k = 1, \ldots, K$, the steering vector is given by

$$\mathbf{a}(\theta) = \left[\begin{array}{cccc} 1 & e^{-j2\pi\frac{d}{\lambda}\sin\theta} & \ldots & e^{-j2\pi(K-1)\frac{d}{\lambda}\sin\theta} \end{array}\right]^T, \tag{1.14}$$

where $(\cdot)^T$ denotes transpose and $\lambda = 2\pi c/\omega_c$ is the wavelength.

The simplest attempt to find the DOA of a source is through conventional *beamforming*. In this approach, a beam is steered in one direction at a time by applying progressive phase shifts to the sensors outputs in order to compensate for the propagation delays of the incident wavefront. The *array beampattern*, $B(\theta_1, \theta_2)$, is defined as the complex gain of a beamformer, when steered in a direction θ_1, to an input plane wave from direction θ_2, i.e.

$$B(\theta_1, \theta_2) = \mathbf{a}^H(\theta_1)\mathbf{a}(\theta_2), \tag{1.15}$$

where $(\cdot)^H$ denotes complex conjugate transpose. Note that for a ULA, (1.15) corresponds to the Discrete Fourier Transform (DFT) of the sequence $a_k(\theta_2)$,

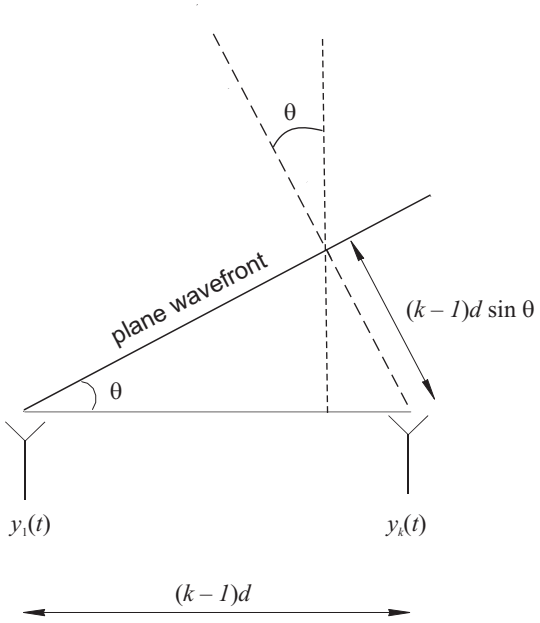

Figure 1.4: A plane wavefront impinging on a ULA.

evaluated at the *spatial frequency* $\phi_1 = 2\pi\frac{d}{\lambda}\sin\theta_1$. As an example, Figure 1.5 shows the magnitude of the beampattern for a ULA with 12 sensors when the array is steered to boresight. Apart from the $\sin\theta$-dependence, the plot resembles the periodogram for estimating a temporal frequency from time series data. The beampattern exhibits a mainlobe at boresight and multiple sidelobes. The width of the mainlobe provides a measure of the ability of the array to resolve two different planewaves. Actually, the beampattern has a profound impact on the achievable performance of any DOA estimation algorithm, see [Mes92].

A common definition of resolution is that the mainlobe peak of one source falls at or outside the first null in the beampattern of the other source. This is referred to as the *Rayleigh resolution limit*, and is thus given by half the null-to-null beamwidth. This resolution limit is with respect to conventional beamforming. This does not mean that two point sources within a beamwidth cannot be resolved with model-based techniques. DOA estimation algorithms that are able to resolve point sources within a beamwidth are therefore sometimes called *superresolution* techniques. Another common definition of beamwidth is the half-power or 3 dB beamwidth, which is defined as the width between the points where the squared magnitude of the beampattern has dropped to one half of its maximum value.

For a ULA to be unambiguous, it can be shown that the sensor spacing

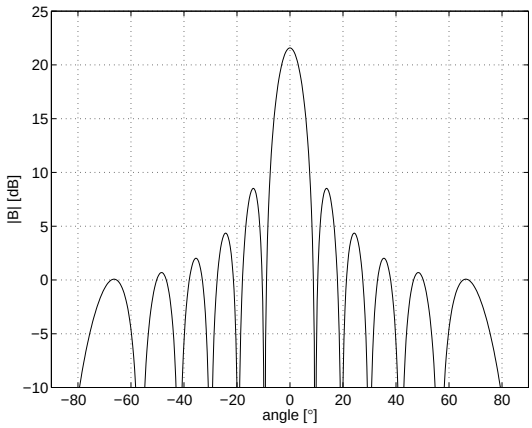

Figure 1.5: Magnitude in dB of the array beampattern for a ULA with 12 sensors, $d = \lambda/2$.

d must be less than $\lambda/2$ if $\theta \in [-\pi/2, \pi/2]$. If $d \geq \lambda/2$ spatial aliasing will occur. This will manifest as additional lobes with the same height as the main lobe. These lobes are called *grating lobes*, and they introduce ambiguities in terms of the position of the maximum peak in the beampattern. Figure 1.6 shows the beampattern for a ULA when the sensor spacing is $d = 2\lambda$. Clearly, even in this noise-free case, it is impossible to tell if the true DOA is $-90°, -30°, 0°, 30°$ or $90°$. However, by employing nonuniform sampling the grating lobes can be suppressed. This is the topic of the next section.

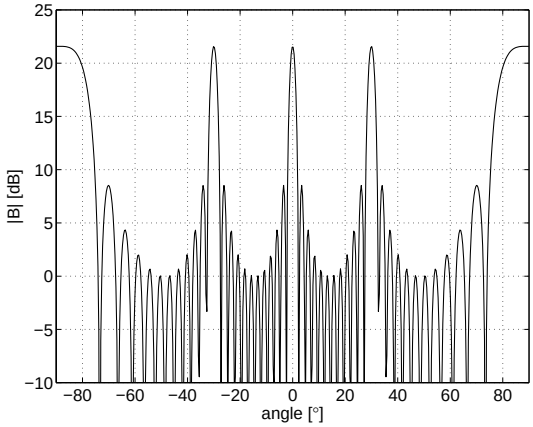

Figure 1.6: Magnitude in dB of the array beampattern for a ULA with 12 sensors, $d = 2\lambda$.

1.3.2 Sparse Arrays

The achievable DOA estimation accuracy depends to a great extent on the array aperture. Large arrays can provide very accurate estimates. However, a large fully populated array is expensive to implement, both in terms of receiver hardware and computational complexity. A way to reduce the cost while maintaining accurate DOA estimates is to reduce the number of antenna elements without decreasing the array size. Such arrays are referred to as *sparse arrays*. The price paid for the reduction of the number of elements is near-ambiguities in the array beampattern, which may lead to large estimation errors. By using nonuniform spacing the ambiguity problems can be reduced. The challenge in sparse array design is then to devise methods for finding array structures that yield high accuracy and yet are robust to ambiguity errors. The research on sparse arrays has been active since the 1950's. See [LL66] for a summary of the early work in this area.

Sparse arrays may be classified into *thinned arrays* and *random arrays*. When designing thinned arrays, the array elements are first placed on a regular grid e.g. a ULA. The sparse array is then obtained by removing a certain number of the elements. In random arrays, the element positions are more or less random. The randomness can be a part of the design procedure or it can be the reality of the application. An example of the latter is when sonobuoys are dropped into the water from an airplane with little control over where they land.

A particular class of thinned arrays that has received a lot of attention in the literature is the so called *minimum redundancy array* (MRA) [Mof68]. Most estimation algorithms are based on exploiting the structure of the spatial covariance matrix. For a ULA with $\lambda/2$ element spacing this matrix contains the spatial covariance lags of all multiples (i.e. $0, 1, \ldots, K-1$) of $\lambda/2$. However, some lags occur many times in the covariance matrix. For instance, the lag $\lambda/2$ is obtained from the element pairs $0 \cdot 1, 1 \cdot 2, \ldots, (K-2) \cdot (K-1)$. The idea behind the minimum redundancy array is to minimize the number of element pairs that have the same spatial covariance lag. In order to be able to estimate all lags in the covariance matrix of the corresponding ULA, there must be at least one element pair for each lag. It has been shown in [Bra62] that zero redundancy is possible only for arrays with $K \leq 4$ elements. For larger arrays one has to search for minimum redundancy with some numerical method. For large arrays this becomes very computationally demanding. The element spacings of MRAs for $K \leq 17$ elements have been tabulated in [LST93]. Another class of thinned arrays are the so called *non-redundant arrays*. In these arrays, zero redundancy is obtained by accepting that some spatial covariance lags are missing. This means that the number of element

pairs that equals a particular lag is either one or zero.

In Figure 1.7 we show the normalized beampatterns, i.e. $|B(\theta_1, 0)|/K$, for a 10 element ULA and MRA respectively. The ULA has a $\lambda/2$ element spacing and the element positions of the MRA are given by $\{0 \cdot 1 \cdot 3 \cdot 6 \cdot 13 \cdot 20 \cdot 27 \cdot 31 \cdot 35 \cdot 36\} \times \lambda/2$. Clearly, the MRA has a narrower mainlobe but higher sidelobes than the ULA. These properties of the MRA suggest a high estimation accuracy at high SNR, but a risk of ambiguity errors at low SNR. More examples of MRAs are given in Part II of the thesis.

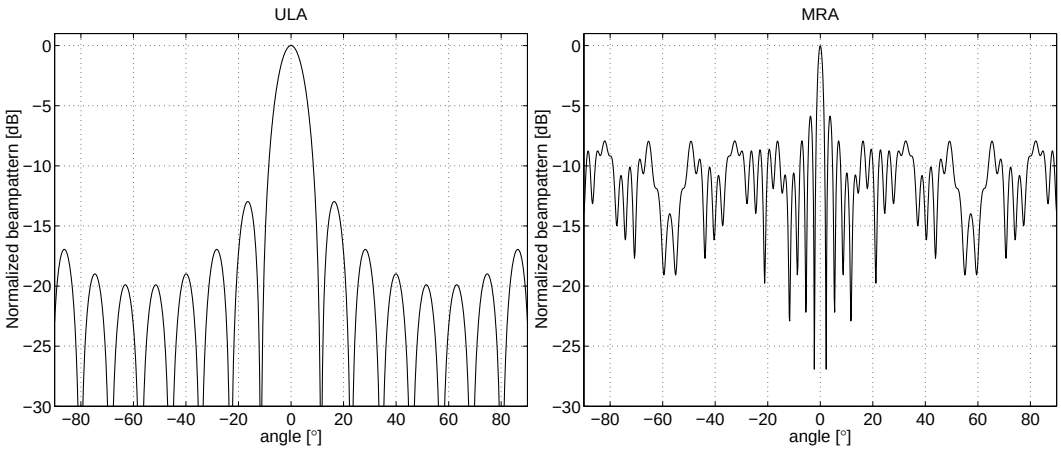

Figure 1.7: Magnitude in dB of the normalized array beampatterns for a ULA (left) and an MRA (right) with 10 elements

1.3.3 Signal and Noise Models

The sensor array model derived so far has been quite general. We will now be more radar specific. The spatial model $\mathbf{a}(\theta)$ will not be discussed in more detail at this stage. Instead, we turn our attention to the temporal model $\mathbf{x}(t)$. Two commonly employed models for $\mathbf{x}(t)$ in the sensor array processing literature are the *stochastic* and *deterministic* models. In the stochastic signal model, $\mathbf{x}(t)$ is modeled as the realization of a stationary, zero mean (usually Gaussian) stochastic process. If the process is modeled as Gaussian, the parameters that need to be determined are the elements in the signal covariance matrix. In the deterministic signal model, each sample of $\mathbf{x}(t)$ is regarded as an unknown, deterministic parameter that needs to be estimated. One may also think of the deterministic model as having conditioned on a particular realization of a random $\mathbf{x}(t)$. Therefore, the models are sometimes also referred to as unconditional and conditional models. A third model,

that will be extensively studied herein, is a *parameterized deterministic* signal model. In this model, each sample $\mathbf{x}(t)$ needs not to be estimated, but the signal is instead described by a relatively small number of parameters. It is sufficient to estimate these parameters in order to completely determine the signal. An example of a parameterized signal model is a known signal waveform with unknown amplitude and phase.

In radar systems modeling, a stochastic signal model is suitable for objects with rapidly fluctuating RCS and for modeling noise jammers. Commonly used models for the statistical distribution of radar objects' RCS are the so called *Swerling cases*, see e.g. [NRC91]. If the RCS is constant during a coherent processing interval, a suitable model is the parameterized deterministic signal model since the complex baseband signal will then be a complex sinusoid. This is due to that when a moving object is illuminated by a radar, the reflected signal return experiences a Doppler shift which is proportional to the relative radial speed between the object and the radar platform. The Doppler shift implies that the complex envelope of the received object signal after matched filtering is a phasor that rotates from pulse to pulse at a rate given by the Doppler frequency. Thus, a suitable model for the return signal at baseband, $x_m(t)$, of the m-th object is a cisoid, $x_m(t) = b_m e^{j\omega_m t}$, with unknown complex amplitude, b_m, (alternatively, real amplitude and phase) and frequency, ω_m. The different time samples $t = t_1, \ldots, t_N$ then correspond to the different pulses for a particular range bin.

We now turn our attention to noise models. There are a number of noise sources in a radar system, e.g. thermal noise generated in the receivers and external noise such as sky noise. Another effect that could be accounted for in the noise term is unwanted echoes from the ground, sea, rain, buildings, etc. This is called *clutter*. The noise will also include any other unmodeled signals, such as intentional jamming. Whereas the thermal noise can be assumed temporally and spatially white, this is not the case for clutter and jamming. If the radar platform is not moving, the clutter return signal will be centered around zero Doppler frequency. It can then be removed relatively easily by a bandstop filter, often called MTI (Moving Target Indication) filter. If the radar platform is moving, the situation is considerably more difficult. The Doppler frequency of the ground echoes will then be different in different directions. One way to deal with this problem is to apply a space-time filter, that removes the clutter. The technique of adaptively suppressing clutter in moving radar systems is often called STAP (Space-Time Adaptive Processing), and has been an intense research area the last decade, see e.g. [War94, Kle98] and the references therein.

One way to deal with colored noise is to use so called *secondary data* to estimate the color. Secondary data are samples which are free from target

signals and only contain noise and interference. The data set which is used in the estimation of the target parameters is referred to as *primary data*. If the secondary data contain noise and interference with the same statistics as the primary data, they can be used to pre-whiten the primary data. This will then bring back the estimation problem to the white noise case. For such a scheme to be successful, a large amount of secondary data may be necessary. In a pulsed Doppler radar, secondary data may be obtained from range bins adjacent to the primary bin under consideration.

In some applications the amount of secondary data may not be sufficient for a successful pre-whitening. For example, due to spatial inhomogeneity of the clutter, there may not be sufficiently many range bins available to obtain an accurate estimate of the spatial covariance matrix. One could then resort to a model of the primary data that accommodates colored noise. The noise parameters then need to be estimated along with the signal parameters, explicitly or implicitly. Sometimes, some structure must be imposed on the correlation of the noise for the estimation problem to be identifiable, see e.g. [SS92].

In this thesis we will adopt the Gaussian model for the noise. While this model is reasonable for thermal noise and some jammers, it is at most only approximately true for clutter. Clutter is often modeled with more heavy-tailed distributions than the Gaussian, such as the Weibull, log-normal and K-distributions (see e.g. [NRC91]). The main reason for using the Gaussian model is simplicity. However, an estimator can also be applied to non-Gaussian data even if it was derived under the Gaussian assumption. Actually, it often turns out that the asymptotic variance of the estimation error depends only on the second order properties (powers and correlations) of the noise, see e.g. [Eri02].

1.4 The Estimation Problem

We now return to the model of the discrete-time received signal vector from a sensor array as given in (1.12):

$$\mathbf{y}(n) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(n) + \mathbf{n}(n), \quad n = 1, \ldots, N. \tag{1.16}$$

It is sometimes convenient to put this model in matrix form according to

$$\mathbf{Y} = \mathbf{A}(\boldsymbol{\theta})\mathbf{X} + \mathbf{N}, \tag{1.17}$$

where

$$\mathbf{Y} = \left[\begin{array}{ccc} \mathbf{y}(1) & \dots & \mathbf{y}(N) \end{array} \right] \tag{1.18}$$

$$\mathbf{X} = \left[\begin{array}{ccc} \mathbf{x}(1) & \dots & \mathbf{x}(N) \end{array} \right] \tag{1.19}$$

$$\mathbf{N} = \left[\begin{array}{ccc} \mathbf{n}(1) & \dots & \mathbf{n}(N) \end{array} \right]. \tag{1.20}$$

The problem of primary interest in this thesis is, given the noisy observations $\mathbf{Y}$, to estimate the DOAs $\boldsymbol{\theta}$. The first part of the thesis is also concerned with estimating the Doppler frequencies, $\boldsymbol{\omega}$, when $\mathbf{x}(n)$ is parameterized according to $\mathbf{x}(n) = \mathbf{B}\mathbf{s}(\boldsymbol{\omega}, n)$, where $\mathbf{B}$ is a diagonal matrix containing the complex amplitudes, and $\mathbf{s}(\boldsymbol{\omega}, n)$ is vector of complex exponentials. However, for ease of exposition, we do not consider this case here in the general introduction.

An immense amount of research literature devoted to this and related problems has been published over the last 30 years. However, it has not been until recently that a textbook that covers most of this area has been published [Tre02]. A more concise overview can be found in [KV96]. Herein, we will only give a brief presentation of the issues that are important for reading this thesis. The interested reader is referred to the above cited references for more details. Before we present some different methods for DOA estimation, we will provide the reader with some basic estimation theory. More details can be found in textbooks on estimation theory and statistical signal processing such as [Tre68, Leh83, Kay93].

1.4.1 Estimation Accuracy

The fundamental question addressed herein is: given the data model in (1.16), how accurately can we estimate θ? For ease of notation, we consider θ to be a scalar parameter at this stage. A desirable property of an estimator, $\hat{\theta}$, of θ is that it is *unbiased*, i.e. that $\mathrm{E}[\hat{\theta}] = \theta$. This is usually a very tough requirement. A weaker property is *consistency* which means that $\hat{\theta} \to \theta$ as $N \to \infty$. If the convergence is "with probability one", the estimator is said to be strongly consistent. If the convergence is "in probability", the estimator is said to be weakly consistent. It can be shown that if an estimator is strongly consistent, it is asymptotically unbiased.

A natural definition of the estimation accuracy is the *mean square error* (MSE) defined as

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{E}\left[(\hat{\theta} - \theta)^2\right], \tag{1.21}$$

where $\mathrm{E}[\cdot]$ denotes expectation. The MSE can be divided into bias and variance according to

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}[\,\hat{\theta}\,] + \mathrm{Bias}^2[\,\hat{\theta}\,]. \tag{1.22}$$

Unfortunately, the bias usually depends on the unknown parameter that is to be estimated, rendering an estimator based on minimization of the MSE usually unrealizable.

A way to study the fundamental limitations in an estimation problem is to study lower bounds on the estimation accuracy. Lower bounds are useful tools for system feasibility studies and evaluation of estimation algorithms. They put fundamental limits on the ultimate estimation performance that can be achieved by any algorithm. In radar and sensor array system design there are many different design parameters that interact with one another. In addition, there are many different criteria to take into account when designing these complex systems, such as cost, accuracy, robustness, etc. These interactions will inevitably lead to trade-offs when designing estimation algorithms. Using performance bounds, an understanding of how different design parameters affect the achievable estimation accuracy can be obtained without having to resort to extensive, time-consuming computer simulations. At best, the dependence of the attainable estimation accuracy upon pertinent design parameters are readily seen by direct inspection of mathematical formulae. In order to arrive at such simple solutions, one often has to make a number of simplifying assumptions. Although these assumptions are not always fulfilled in practice, the bounds can still be expected to lower bound actual performance.

For a lower bound to be useful, it should be as *tight* as possible, i.e. there should exist an estimation algorithm whose accuracy is reasonably close to the bound. If an estimator has been found that achieves a lower bound, there is no use in putting more work trying to find an estimator with better accuracy. If, on the other hand, there is a gap between the performance of an estimator and the lower bound, the situation is more difficult. Then, we cannot tell if there exists a better estimator, we just have not found it, or if a tighter bound can be found.

By far the most used bound is the Cramér-Rao bound (CRB) [Fis22, Rao45, Cra46]. The CRB states that for any unbiased estimator $\hat{\theta}$,

$$\mathrm{Var}[\,\hat{\theta}\,] \geq \frac{1}{-\mathrm{E}\left[\dfrac{\partial^2 \log p(\mathbf{Y};\theta)}{\partial \theta^2}\right]}, \tag{1.23}$$

where $p(\mathbf{Y};\theta)$ is the *likelihood function* of the observed data, and the derivative is evaluated at the true value of θ. The likelihood function is the probability density function (PDF) of $\mathbf{Y}$, when it is viewed as a function of θ for a fixed $\mathbf{Y}$. It is a measure of how likely it is that the observed data $\mathbf{Y}$ were generated by the model with parameter value θ. For the CRB to hold, the PDF must satisfy some regularity conditions, see e.g. [Tre68, Leh83, Kay93].

The denominator in (1.23) is referred to as the *Fisher information* of the data.

For a vector parameter, $\boldsymbol{\theta}$, the CRB is given by

$$\mathrm{E}\left[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\right] \geq \mathbf{J}^{-1}, \tag{1.24}$$

where

$$\{\mathbf{J}(\boldsymbol{\theta})\}_{mn} = -\mathrm{E}\left[\frac{\partial^2 \log p(\mathbf{Y};\theta)}{\partial\theta_m \partial\theta_n}\right] \tag{1.25}$$

and the notation $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A}-\mathbf{B}$ is positive semidefinite. The matrix $\mathbf{J}(\boldsymbol{\theta})$ is called the *Fisher information matrix* (FIM).

An estimator that has an error variance equal to the CRB is called *efficient*. In general, it is hard to find such an estimator. However, it is often possible to find an estimator that attains the CRB asymptotically as $N \to \infty$. Such as estimator is said to be *asymptotically efficient*.

An estimator that is closely related to the CRB is the *maximum likelihood* (ML) estimator. As the name suggests, this estimator is obtained by finding the maximizing argument of the likelihood function, $p(\mathbf{Y};\theta)$. An interpretation of this method is then that it finds the parameter value that most likely generated the observed data. It can be shown that *if* an efficient estimator exists, this estimator is the ML estimator. Furthermore, it can be shown that under certain regularity conditions (see e.g. [Tre68, Leh83, Kay93]), the ML estimator is consistent and asymptotically efficient.

The main reasons for the widespread use of the CRB are that it is relatively easy to compute, and that there exists an estimator that (normally) attains it asymptotically. The CRB is thus an asymptotically tight lower bound. However, for finite samples there is usually no guarantee that the CRB is tight. Actually, it is well known that the ML estimator usually exhibits a so called *threshold effect*. For few samples, or low SNR, the estimation error at some point starts to increase vary rapidly. This effect is not visible in the CRB. Hence, below this threshold, the estimation error variance of the ML estimator is much higher than the CRB.

The reason for this behavior of the ML estimator is that in nonlinear estimation problems, the likelihood function is often multimodal. In the asymptotic region, the estimates are close to the true value of the unknown parameter. This means that they are within the "mainlobe" of the likelihood function. Below the threshold, noise may cause "ambiguity" or "sidelobe" peaks to have higher magnitude than the "mainlobe" peak. This results in a dramatic increase of the estimation error. According to (1.23), the CRB is related to the curvature of the likelihood function at the true parameter

value. An ambiguity peak in the likelihood function far away from the true parameter value has no influence on the CRB. Therefore, the threshold effect is not captured by the CRB.

A natural question to ask is then: do better estimators than the ML estimator exist, or do better bounds than the CRB exist? To the first question there is no general answer, at least to the author's knowledge. The answer to the second question is yes. A number of bounds have been developed over the years that are tighter than the CRB below the threshold. The most popular among these bounds are the Barankin [Bar49], Ziv-Zakai [ZZ69], and Weiss-Weinstein bounds [WW85]. Although these bounds do capture the threshold effect, they do not perfectly predict the estimation error of the ML DOA estimator. In Part II of this thesis, we present an extensive study of the threshold region performance of the ML estimator.

1.4.2 Estimation Methods

In this section, we give a brief overview of different DOA estimation algorithms. More details can be found in e.g. [Tre02].

Beamforming Methods

As mentioned previously, a natural approach to find the DOA of a single source is to scan a beam over the search volume and monitor the received output power. The DOA estimate is then taken as the pointing direction that corresponds to the peak output power. The beam is formed by compensating for the time delay between different sensors that a plane wave experiences as it travels across the array. If the source signal is narrowband, the time delay amounts to a simple phase shift. Mathematically, the beamforming can be expressed as a multiplication of the received signal vector with a weight vector $\mathbf{w}$ according to

$$\mathbf{z}(t) \;=\; \mathbf{w}^H(\theta)\mathbf{y}(t), \quad t = 1, \ldots, N. \tag{1.26}$$

The output power, averaged over the N snapshots, is then given by

$$P(\theta) \;=\; \frac{1}{N}\sum_{t=1}^{N} |\mathbf{w}^H(\theta)\mathbf{y}(t)|^2 = \mathbf{w}^H(\theta)\hat{\mathbf{R}}\mathbf{w}(\theta), \tag{1.27}$$

where

$$\hat{\mathbf{R}} \;=\; \frac{1}{N}\sum_{t=1}^{N} \mathbf{y}(t)\mathbf{y}^H(t) \tag{1.28}$$

is the *sample covariance matrix*. Depending on how the weight vector $\mathbf{w}$ is determined, different beamforming methods result. Here, we will present two commonly used methods: the conventional and the Capon beamformers. A more extensive treatment of various beamforming methods can be found in [VB88].

Conventional Beamformer In the conventional beamforming method, the weight vector is chosen so as to maximize the output power when the beam is steered to the true direction, θ. This is a spatially matched filter, for which the solution in spatially white noise is given by

$$\mathbf{w}_{\mathrm{BF}}(\theta) = \frac{\mathbf{a}(\theta)}{|\mathbf{a}(\theta)|}. \tag{1.29}$$

Since the true direction is unknown, a number of hypothesized directions are evaluated, and the one for which the output power is the maximum is taken as the DOA estimate. This scheme amounts to finding the maximum in the beamformer spectrum, P_{BF}, which is obtained by inserting (1.29) into (1.27)

$$P_{\mathrm{BF}}(\theta) = \frac{\mathbf{a}^H(\theta)\hat{\mathbf{R}}\mathbf{a}(\theta)}{\mathbf{a}^H(\theta)\mathbf{a}(\theta)}. \tag{1.30}$$

This spectrum is a spatial analog of the Fourier based spectral methods used in times series analysis [SM97]. In the case of a ULA with a steering vector according to (1.14) the beamformer spectrum in (1.30), when viewed as a function of $\phi = 2\pi d \sin(\theta)/\lambda$, is equivalent to the classical periodogram. This means that the conventional beamformer also suffers from the limited resolution of Fourier methods. For the DOA estimation problem, this means that the resolution of the conventional beamformer cannot be better than the array beamwidth, no matter how many snapshots we collect or how high the SNR is.

Capon's Beamformer The limited resolution of the conventional beamformer is essentially due to spectral leakage between closely spaced sources. A way to alleviate this problem was proposed by Capon in [Cap69]. The idea behind this approach is to minimize the output power from the beamformer, while protecting the gain in the look direction, i.e.

$$\begin{aligned} &\min_{\mathbf{w}} P(\mathbf{w}) \\ &\text{subject to} \quad \mathbf{w}^H\mathbf{a}(\theta) = 1. \end{aligned} \tag{1.31}$$

This method is also frequently referred to as the minimum variance distortionless response (MVDR) beamformer. The solution to the constrained

optimization problem in (1.31) is readily obtained using the method of Lagrange multipliers. The result is

$$\mathbf{w}_{\mathrm{CAP}} = \frac{\hat{\mathbf{R}}^{-1}\mathbf{a}(\theta)}{\mathbf{a}^H(\theta)\hat{\mathbf{R}}^{-1}\mathbf{a}(\theta)}, \tag{1.32}$$

which yields the Capon spectrum according to

$$P_{\mathrm{CAP}}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\hat{\mathbf{R}}^{-1}\mathbf{a}(\theta)}. \tag{1.33}$$

The Capon beamformer tries to null any source that is not in the look direction. This results in an improved resolution as compared to the conventional beamformer. The resolution depends on SNR, but not (much) on the number of snapshots. However, the resolution capability of the Capon beamformer is still rather poor.

Maximum Likelihood Methods

The ML method is a standard technique in statistical estimation theory. In order to apply the maximum likelihood method, the likelihood function of the observed data needs to be determined. Depending on how the source signals are modeled, two main ML methods have been developed for the DOA estimation problem. If the source signals are modeled as Gaussian stochastic processes, we obtain the stochastic ML (SML) method. If the signals are considered as deterministic parameters that need to be estimated along with the DOAs, the so-called deterministic ML (DML) estimator results.

The Stochastic Maximum Likelihood Method This method is based on the assumption that the source signals are zero-mean, complex Gaussian stochastic processes. The information about the source parameters is then contained in the covariance matrix of the received data. More specifically, the signals, $\mathbf{x}(n)$, and noise, $\mathbf{n}(n)$, are assumed to be independent, zero-mean, complex Gaussian random processes with second-order moments

$$\begin{aligned} \mathrm{E}\left[\mathbf{x}(m)\mathbf{x}^H(n)\right] &= \mathbf{P}\delta_{mn} \qquad \mathrm{E}\left[\mathbf{n}(m)\mathbf{n}^H(n)\right] = \sigma^2\mathbf{I}\delta_{mn} \\ \mathrm{E}\left[\mathbf{x}(m)\mathbf{x}^T(n)\right] &= \mathbf{0} \qquad\quad \mathrm{E}\left[\mathbf{n}(m)\mathbf{n}^T(n)\right] = \mathbf{0}, \end{aligned} \tag{1.34}$$

where $\mathbf{P}$ is the unknown signal covariance matrix and δ_{mn} represents the Kronecker delta. Thus, the signals are assumed to be temporally white, and the noise is assumed to be spatially and temporally white. From (1.16) and

the assumptions stated above we obtain the data covariance matrix according to

$$\begin{aligned} \mathrm{E}\left[\mathbf{y}(m)\mathbf{y}^H(n)\right] &= \mathbf{R}\,\delta_{mn} = \left(\mathbf{A}(\boldsymbol{\theta})\mathbf{P}\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I}\right)\delta_{mn} \quad (1.35)\\ \mathrm{E}\left[\mathbf{y}(m)\mathbf{y}^T(n)\right] &= \mathbf{0}. \quad (1.36)\end{aligned}$$

The unknown parameters in this model are the DOAs in $\boldsymbol{\theta}$, the elements in the signal covariance matrix, $\mathbf{P}$, and the noise variance, σ^2. With the assumptions above, the likelihood function of the observed data will be that of the complex K-variate Gaussian distribution, see e.g. [Goo63]. The likelihood function of a single snapshot is then given by

$$p(\mathbf{y};\boldsymbol{\theta},\mathbf{P},\sigma^2) = \frac{1}{\pi^K|\mathbf{R}|}\exp\left[-\mathbf{y}^H\mathbf{R}^{-1}\mathbf{y}\right], \quad (1.37)$$

where $|\cdot|$ denotes the determinant of a matrix. Since the snapshots are independent and identically distributed, the likelihood function of all snapshots in $\mathbf{Y}$ is given by

$$p(\mathbf{Y};\boldsymbol{\theta},\mathbf{P},\sigma^2) = \prod_{n=1}^{N}\frac{1}{\pi^K|\mathbf{R}|}\exp\left[-\mathbf{y}^H(n)\mathbf{R}^{-1}\mathbf{y}(n)\right]. \quad (1.38)$$

Maximizing the likelihood function is equivalent to minimizing the negative log-likelihood function, given by

$$-\log[p(\mathbf{Y};\boldsymbol{\theta},\mathbf{P},\sigma^2)] = KN\log\pi + N\log|\mathbf{R}| + \sum_{n=1}^{N}\mathbf{y}^H(n)\mathbf{R}^{-1}\mathbf{y}(n). \quad (1.39)$$

The stochastic ML estimator is then obtained solving the following optimization problem (ignoring the constant term and normalizing by N)

$$[\hat{\boldsymbol{\theta}},\hat{\mathbf{P}},\hat{\sigma}^2] = \arg\min_{\boldsymbol{\theta},\boldsymbol{P},\sigma^2} V_{\mathrm{SML}}(\boldsymbol{\theta},\mathbf{P},\sigma^2), \quad (1.40)$$

where

$$\begin{aligned} V_{\mathrm{SML}}(\boldsymbol{\theta},\mathbf{P},\sigma^2) &= \log|\mathbf{R}| + \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}^H(n)\mathbf{R}^{-1}\mathbf{y}(n)\\ &= \log|\mathbf{R}| + \mathrm{Tr}\Big\{\mathbf{R}^{-1}\hat{\mathbf{R}}\Big\}. \quad (1.41)\end{aligned}$$

If no other structure is imposed on $\mathbf{P}$ than being Hermitian, the total number of parameters to estimate in this problem is $M^2 + M + 1$. Since this

optimization problem is highly nonlinear, it becomes computationally very demanding when $M > 1$. The problem can be simplified by concentrating the ML criterion function with respect to $\mathbf{P}$ and σ^2. This means that only a search over the unknown DOA parameters is necessary; thereby reducing the dimension of the optimization problem to M. The concentrated form is given by [Böh86, SN95]

$$V_{\text{SML}}(\boldsymbol{\theta}) = \log \left| \mathbf{A}(\boldsymbol{\theta}) \hat{\mathbf{P}}(\boldsymbol{\theta}) \mathbf{A}^H(\boldsymbol{\theta}) + \hat{\sigma}^2(\boldsymbol{\theta}) \mathbf{I} \right|, \tag{1.42}$$

where

$$\hat{\mathbf{P}}(\boldsymbol{\theta}) = \mathbf{A}^{\dagger}(\boldsymbol{\theta}) \big[\hat{\mathbf{R}} - \hat{\sigma}^2(\boldsymbol{\theta}) \mathbf{I} \big] \mathbf{A}^{\dagger H}(\boldsymbol{\theta}), \tag{1.43}$$

$$\hat{\sigma}^2(\boldsymbol{\theta}) = \frac{1}{K-M} \text{Tr} \Big\{ \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) \hat{\mathbf{R}} \Big\}. \tag{1.44}$$

Here, $\mathbf{A}^{\dagger}$ is the pseudo-inverse of $\mathbf{A}$ and $\mathbf{P}_{\mathbf{A}}^{\perp}$ is the orthogonal projection onto the null space of $\mathbf{A}^H$, i.e.

$$\mathbf{A}^{\dagger} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \tag{1.45}$$

$$\mathbf{P}_{\mathbf{A}}^{\perp} = \mathbf{I} - \mathbf{A}\mathbf{A}^{\dagger}. \tag{1.46}$$

Even though the SML method was derived under the assumption of Gaussian data, it can certainly be applied to non-Gaussian data as well. An interesting property is that the asymptotic accuracy of the SML estimator depends only on the second-order (powers and correlations) of the signals, irrespective of their actual distribution [SN90b, OVK92].

The Deterministic Maximum Likelihood Method In the DML method, no assumptions on the distribution of the signals are made. Each sample of the signals is instead considered as an unknown parameter that needs to be estimated. The noise is still considered to be a zero-mean, complex Gaussian random process. The information about the source parameters is then contained in the mean of the received data. The first- and second-order moments are in this case given by

$$\text{E}\left[\mathbf{y}(n)\right] = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(n) \tag{1.47}$$

$$\text{E}\left[(\mathbf{y}(m) - \text{E}\left[\mathbf{y}(m)\right])(\mathbf{y}(n) - \text{E}\left[\mathbf{y}(n)\right])^H \right] = \sigma^2 \mathbf{I} \delta_{mn} \tag{1.48}$$

$$\text{E}\left[(\mathbf{y}(m) - \text{E}\left[\mathbf{y}(m)\right])(\mathbf{y}(n) - \text{E}\left[\mathbf{y}(n)\right])^T \right] = \mathbf{0}. \tag{1.49}$$

The unknown parameters in this model are the DOAs, $\boldsymbol{\theta}$, the signal samples in $\mathbf{X}$, and the noise variance, σ^2. The PDF of the N snapshots is given by

$$p(\mathbf{Y};\boldsymbol{\theta},\mathbf{X},\sigma^2) = \prod_{n=1}^{N} \frac{1}{|\pi\sigma^2\mathbf{I}|} \exp\left[-\sigma^{-2}\{\mathbf{y}(n)-\mathbf{A}(\boldsymbol{\theta})\mathbf{x}(n)\}^H\{\mathbf{y}(n)-\mathbf{A}(\boldsymbol{\theta})\mathbf{x}(n)\}\right]. \tag{1.50}$$

The negative log-likelihood function then takes the form

$$-\log[p(\mathbf{Y};\boldsymbol{\theta},\mathbf{X},\sigma^2)] \;=\; KN\log(\pi\sigma^2) + \frac{1}{\sigma^2}\|\mathbf{Y}-\mathbf{A}(\boldsymbol{\theta})\mathbf{X}\|_F^2, \tag{1.51}$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix. The DML criterion function can also be concentrated with respect to the signal and noise parameters. This results in that the DML DOA estimator is given by [Böh84, Wax85]

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} V_{\mathrm{DML}}(\boldsymbol{\theta}), \tag{1.52}$$

where

$$V_{\mathrm{DML}}(\boldsymbol{\theta}) = \mathrm{Tr}\Big\{\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\hat{\mathbf{R}}\Big\}. \tag{1.53}$$

Hence, the ML estimator under the deterministic signal model is also obtained by finding the minimum of a multidimensional, nonlinear function. Both the SML and DML estimators have excellent performance, but they are computationally demanding. This motivates the need for estimators having performance and computational complexity between those of the beamforming and ML methods. One such class of estimators is briefly described in the following.

Subspace Methods

A class of estimators that has become very popular are the so-called *subspace methods*. These methods have very good performance at a reasonable computational cost. The great interest was sparked by the introduction of the MUSIC (Multiple Signal Classification) algorithm in [Sch79]. The subspace methods rest on a geometrical interpretation of the data model in (1.16). If the number of signals is less than the number of sensors, the signal part will reside in subspace of the full complex K-dimensional vector space. In the MUSIC algorithm, a pseudo-spectrum is formed that is a measure of the distance between the array manifold and the orthogonal complement of

the signal subspace, called the noise subspace. The DOA estimates are then obtained by locating the peaks in this pseudo-spectrum.

Another popular subspace method that is only applicable to certain array geometries is the ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) algorithm [RPK86]. Later, it was discovered that many methods can be cast in a common subspace fitting framework [VO91]. Among these are the beamforming, DML, MUSIC, and ESPRIT methods. Based on these insights, a subspace method called WSF (Weighted Subspace Fitting) was derived that has the same asymptotic accuracy as the SML method, but at a lower computational cost.

Since we do not deal with subspace methods in this thesis, no details are given here. The interested reader is referred to the above cited references for further reading.

Estimating the Number of Sources

All methods considered herein, except the beamforming methods, require knowledge of the number of sources. Since this knowledge is normally not available a priori, this number has to be estimated. Common methods when using subspace methods are the AIC (An Information Criterion) and MDL (Minimum Description Length) methods[2] [WK85]. These methods are separated from the DOA estimation problem. Another approach is to jointly estimate the number of sources and their DOAs. Example of such approaches are given in [WZ89, Wax91, OVSN93]. We do not deal with the problem of estimating the number of sources in this thesis, but assume, whenever required, that this number is known.

1.5 Motivation and Aim of the Thesis

Model-based signal processing and digital arrays have the potential of significantly improving the performance of modern radar systems. However, this comes at the expense of increased costs and complexity, both in terms of hardware and computational requirements. Furthermore, the performance of model-based methods are sensitive to modeling errors, requiring accurate system calibration. These facts have severely limited the introduction of these techniques into real world applications. Reduction of hardware and computational complexity is therefore an issue of utmost importance.

[2]More precisely, these methods estimate the signal subspace dimension which is equal to the number of non-coherent sources

Two different aspects of reducing the complexity have been the motivation behind the work presented in this thesis:

- Reducing computational complexity by simplifying estimation algorithms. This is the motivation behind Part I of this thesis.
- Reducing hardware (and computational) complexity by reducing the number of antenna elements. This is the motivation behind Part II of this thesis.

The focus of this thesis is on the fundamental limitations on the achievable performance rather than on proposing algorithms for practical implementation. To this end, we study lower bounds on the achievable estimation accuracy and the theoretical performance of the best known estimators. The results could be useful for feasibility studies and in the early stages of system design. Although practical aspects such as algorithm design and hardware implementation issues are not the primary focus, they are indeed motivations behind the work presented herein.

In the first part of the thesis it is demonstrated how the DOA and Doppler estimation algorithms can be simplified by adopting a decoupled approach. The second part has its origin from demands on reducing hardware complexity by reducing the number of antenna elements. In order to maintain accurate DOA estimation, this reduction is performed without decreasing the array aperture.

These types of complexity reductions may have a substantial influence on the estimation performance. It is therefore important to have adequate tools to analyze these types of complexity/performance trade-offs. The focus of this thesis is on performance analysis pertaining to the two types of complexity reductions mentioned above. The first part of the thesis is concerned with performance analysis of joint and decoupled DOA/Doppler estimation. In the second part a performance analysis of DOA estimation is presented that is focused on incorporating the effect of ambiguity errors into the analysis. This is of particular importance when dealing with sparse arrays, since an intrinsic property of such arrays are high sidelobes or near-ambiguities. These high sidelobes may cause large estimation errors which set limits to the degree of sparsity that can be tolerated.

1.6 Outline and Contributions

The thesis is divided into two self-contained parts. Both parts are concerned with the achievable performance of parameter estimators in sensor array signal processing. The first part is about joint and decoupled angle-frequency

estimation, whereas the second deals only with the angle estimation problem. These problems can be viewed as 2-D and 1-D frequency estimation problems, respectively. However, the major difference between the two parts is in the focus of the performance analysis. The results in Part I are mainly of asymptotic nature, i.e. the amount of data is assumed to be large. In this asymptotic region conventional performance analysis techniques such as Taylor expansions and CRBs are useful tools. In Part I CRBs and asymptotic analyses of the maximum likelihood estimators are developed for the angle-frequency estimation problem.

In many practical situations the amount of data may not be sufficient for the asymptotic assumptions to be valid. In such scenarios asymptotic analyses usually give too optimistic performance predictions. It is therefore of great interest to develop tools that can be used to predict the performance also in the non-asymptotic region. A prominent feature of non-linear estimation in the non-asymptotic region is the threshold effect. In part II of this thesis a performance analysis of the maximum likelihood DOA estimator that takes the threshold effect into account is developed. Finding the finite sample properties of non-linear estimators has generally been considered to be a difficult problem. Yet, it is shown in this thesis that an analysis based on very simple ideas can be used to predict the non-asymptotic performance of the ML DOA estimator with high accuracy.

The main contributions of this thesis are summarized chapter-wise in the following outline. Edited versions of Chapters 2 – 7 will be published as one chapter in the book [HGC].

Chapter 2

The first chapter in Part I gives a brief background to the problem of angle-frequency estimation and a review of the research literature in the area. Further, it includes problem formulation and a discussion on the assumptions that we make. The chapter ends by showing the close connection to 2-D frequency estimation.

Chapter 3

In this chapter joint angle-frequency estimation using a 2-D model is considered. The results presented in Chapter 3 and 4 are extensions of related work that has been previously reported by other authors. The extensions are to multiple sources, more general noise models, and relaxed restrictions on sampling and asymptotics.

Firstly, a general CRB for parameterized signals in colored noise is presented. This CRB is then used to derive the CRB for joint angle-frequency estimation using the 2-D model. Asymptotic CRBs are then derived, assuming the amount of data is large in one of the frequency dimensions. It is shown that the multi-source CRB in this case decouples to the single source CRB. Maximum likelihood and weighted least squares techniques for joint angle-frequency estimation are then reviewed. It is shown that these estimators have excellent performance but are computationally very expensive.

Chapter 4

Recognizing the computational complexity of the 2-D estimators, a decoupled 1-D/1-D model is studied in Chapter 4. Firstly, the finite and asymptotic CRBs for the 1-D/1-D model are presented. A decoupled weighted least-squares (WLS) estimator based on the 1-D/1-D model is then proposed. It is shown that the multi-source 1-D/1-D WLS estimator can achieve the single source 2-D CRB asymptotically. It is sufficient that the amount of data is large in only one of the frequency dimensions for the estimator to achieve the 2-D CRB in both dimensions. In order to achieve this, the estimation must be performed in the right order. The frequency dimension in which the amount of data is large must be processed first. In this dimension, the sources are resolvable and, subsequently, the estimation in the other frequency dimension can be performed for one source at a time. The result is a tremendous reduction in computational complexity as compared to the full 2-D frequency estimation problem. Yet, under the stated assumptions, the performance loss is insignificant. This result is the main contribution of Chapter 4.

Chapter 5

The methods presented in Chapter 3 and 4 are all implicit in nature. The parameter estimates are defined as the minimizing arguments of some non-linear criterion function. In practice, finding the minimum is far from trivial, especially when the number of parameters is large. The presented objective functions are all of the nonlinear least squares type. In Chapter 5 a computationally attractive and reasonably reliable procedure to solve such problems is described. The technique is based on local numerical optimization, initialized using the so-called RELAX procedure.

Chapter 6

Chapter 6 concludes the first part of the thesis by presenting some numerical examples and simulation results in order to corroborate the theoretical analysis presented in the preceding chapters.

Chapter 7

Chapter 7 concludes the first part of the thesis by giving a summary of the main findings in Chapter 3 – Chapter 6.

Chapter 8

In the second part of the thesis the threshold region performance of the maximum likelihood DOA estimator is analyzed. Both the deterministic and stochastic maximum likelihood estimators are considered. Chapter 8 provides an introduction to Part II by explaining the threshold effect and reviewing the research literature in the area. Further, the contributions presented in this thesis are compared with previous and present-day work by other authors.

Chapter 9

For the case of a single signal in white Gaussian noise, the ML estimator is equivalent to the conventional beamformer. This case is studied in Chapter 9. Approximations to the probability of outlier and MSE are derived. It is verified by simulations that the derived approximations provide an accurate characterization of the threshold region performance of the ML estimators. It is also shown that for a single snapshot the ML estimator under the stochastic signal model is not statistically efficient as the SNR tends to infinity due to the effect of outliers.

Chapter 10

The results in Chapter 9 are extended in Chapter 10 to multiple sources. The basic ideas behind the approximations of the MSE and probability of outlier remain the same but the technical details become considerably more intricate. Finding the desired probabilities amounts to computing tail probabilities of the distribution of quadratic forms.

The main contribution of Chapter 9 and 10 is that together they provide a comprehensive treatment of the threshold region performance of ML DOA estimation. Although similar ideas and analyses have appeared previously for

simpler cases and other applications, to the author's knowledge, no extensive study of the problem considered herein has been published before. Several authors have used bounding techniques to characterize the threshold region performance in DOA estimation. Indeed, these techniques do set a limit to the ultimate accuracy that can be achieved, but often the bounds are not tight i.e. no estimator that reaches these bounds is known. It can therefore be argued that the performance of the ML estimator is, in a sense, of more practical interest since it is often used as a benchmark estimator.

Chapter 11

The second part of the thesis ends with two application examples. In the first example the single-source performance of an interferometric-like DOA estimation system consisting of two widely separated ULAs is investigated. Several authors have treated various, more narrow, aspects of this type of array structure. The purpose with this example is to explore further the fundamental limitations in DOA estimation with this particular array structure. It includes a discussion on optimal Bayesian estimation and performance bounds and how they relate to their classical counterparts. The example provides some further insights into estimation problems that are prone to ambiguities.

In the second example a novel criterion for optimizing the element positions in sparse arrays is proposed. Previous approaches to this problem include minimization of the CRB and optimization of various antenna pattern characteristics. A problem with the CRB approach is that it is a local measure that does not capture the effect of near-ambiguities. A deficiency with the antenna pattern approach is that the connection to the resulting estimation accuracy may not always be clear. Herein, the threshold in the Weiss-Weinstein lower bound is used as optimization criterion. This criterion has the desirable property that it is a measure of achievable estimation accuracy taking the threshold effect into account. An interesting result of the optimization is that the optimal array has a nearly identical lower bound to the so called minimum redundancy array, which was devised from completely different ideas.

Chapter 12

Chapter 12 ends the thesis by summarizing the main results of Part II.

1.6.1 Publications

The thesis is based on the following publications by the author:

Book Chapter

- F. Athley, M. Viberg, and J. Eriksson. "High-Resolution Space-Time Signal Processing for Radar", In Y. Hua, A. Gershman, Q. Cheng, editors, *High-Resolution and Robust Signal Processing*, Dekker, New York, NY, To be published.

Conference Papers

- F. Athley. "Threshold Region Performance of Deterministic Maximum Likelihood DOA Estimation of Multiple Sources". In *Proc. 36th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Nov. 2002.
- F. Athley. "Performance Analysis of DOA Estimation in the Threshold Region". In *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Processing*, Orlando, FL, USA, May 2002.
- F. Athley. "Threshold Region Performance of Maximum Likelihood DOA Estimation For a Single Source". In *Proc. 10th Annual Adaptive Sensor Array Processing Workshop*, MIT Lincoln Lab., Lexington, MA, USA, March 2002.
- F. Athley. "Optimization of Element Positions for Direction Finding with Sparse Arrays". In *Proc. 11th IEEE Signal Processing Workshop on Statistical Signal Processing*, Singapore, Aug 2001.
- F. Athley and C. Engdahl. "Direction-Of-Arrival Estimation Using Separated Subarrays". In *Proc. 34th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Oct 2000.
- F. Athley. "Asymptotically Decoupled Angle-Frequency Estimation with Sensor Arrays". In *Proc. 33rd Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Oct 1999.

Technical Report

- F. Athley, "Angle and Frequency Estimation Using Sensor Array", Technical report 370L, Licentiate thesis, Chalmers University of Technology, Feb 2001.

1.7 Topics for Future Research

We end this chapter with some suggestions for future research topics that are intimately related to the material presented in the thesis.

Noise Models

The noise models employed in this thesis may not be realistic in all applications. In Part I, the space-time noise covariance matrix is assumed to have a Kronecker structure. As discussed in Chapter 2, this may not be a realistic assumption in a scenario involving clutter and jamming. A topic for future research is to investigate if this imposed structure can be relaxed. A recent contribution to this area is [MS02], where the asymptotic CRB for 2-D frequency estimation is derived assuming a block Toeplitz structure on the space-time noise covariance matrix. Another recent contribution is [CF02], where an asymptotic analysis of least squares estimation of real 2-D sinusoids in colored noise is performed.

One of the most severe interference sources in a radar system is clutter. Clutter has a rich structure and an interesting research topic is to develop structured models for the clutter that could be used in the estimation methods presented herein. This would alleviate the requirements on a large amount of secondary data.

In Part II, spatio-temporally white noise is assumed throughout. It would be interesting to extend the results in Part II to colored noise. In [Ric03] the case of a single deterministic signal in white noise is extended to the case of colored noise and access to secondary data.

Threshold Region Performance of Angle-Frequency Estimation

A natural extension of the work presented in this thesis is to apply the threshold region analysis in Part II to the estimation problem in Part I.

Lower Bounds on Estimation Accuracy

It would be nice to bridge the gap between the MSE approximation of the ML estimator derived in Part II and the lower bounds available today. It is unclear if tighter bounds can be found or if estimators with better threshold region performance than the ML estimator exist for the multi-source problem. The connection between the MSE approximation and the Weiss-Weinstein and Ziv-Zakai lower bounds should be established. Maybe this could provide new insights that lead to improved bounds. Bayesian bounds for the multi-source problem is also a topic for future research.

Optimization of Element Positions in Sparse Arrays

The MSE approximation could be used to optimize the element positions in sparse arrays in a similar way that we used the Weiss-Weinstein bound in Chapter 11. This has the potential of being a faster and more accurate method. More importantly, it could be applied to the multi-source problem. Instead of optimizing the accuracy for a single source, we could e.g. optimize the resolution capability of the array.

Further Analysis of the Stochastic ML DOA Estimator

We show in Part II that the SML estimator is statistically inefficient for high SNR in the case of a single signal and a single snapshot. Simulation results indicate that this observation also holds for multiple signals and multiple snapshots. However, we have not verified this theoretically. To our knowledge, no high SNR analysis of the SML estimator, when applied to stochastic signals, has been reported in the literature[3].

The analysis of the threshold region performance of the SML estimator in the case of multiple sources is not complete. The distribution of the concentrated SML criterion function remains to be computed. It would also be desired to find a way to concentrate the SML criterion so that positive semi-definite estimates of the signal covariance matrix can be guaranteed. How to treat the nuisance parameters in the MSE approximation for the case of correlated signals also deserves some more attention.

[3]In [Vib93] a high SNR analysis of the SML estimator when applied to a deterministic signal model was performed.

Part I
Asymptotically Decoupled Angle-Frequency Estimation with Sensor Arrays

Chapter 2

Introduction

2.1 Background and Overview

The first part of this thesis is concerned with the problem of estimating the directions of arrival (DOAs) and Doppler frequencies of multiple targets with a pulsed Doppler radar employing an array antenna. This can also be viewed as a two-dimensional (2-D) frequency estimation problem. Estimating the frequencies of 2-D sinusoids has received considerable interest in the literature, see e.g. [Kay80, McC82, DM84, Hua92, RZZ94, LvdVD98, Eri02]. This problem finds applications in many fields such as sonar, radar, geophysics, radio astronomy, radio communications, and medical imaging. The material herein is presented in the context of pulsed Doppler radar, although the results are generally applicable to separable 2-D data models.

We consider the case where an array of spatially distributed sensors receives sinusoidal return signals from distant sources. Our main objective herein is to estimate the temporal and spatial frequencies of the source signals. The temporal frequency is related to the relative speed between a source and the radar platform via the well known Doppler effect. The spatial frequency is related to the source's direction of arrival via the inter-element phase shifts that are present when a narrowband wavefront is incident upon the array.

The chief systematic approach to solving parameter estimation problems is the maximum likelihood (ML) method. Under fairly general assumptions, the ML approach is known to produce asymptotically unbiased and statistically efficient estimates, see e.g. [Kay93]. The latter means that the ML estimates asymptotically achieve the Cramér-Rao bound (CRB), which is a lower bound on the error variance of any unbiased estimator. The CRB for the full 2-D data model is derived in Chapter 3, followed by a simplified

expression, which is valid for a large number data samples (either temporally or spatially). The exact ML estimator (MLE) is also derived, as well as an asymptotically efficient approximation based on nonlinear weighted least squares (WLS). Unfortunately, for the problem at hand, the ML and WLS approaches result in multidimensional, nonlinear searches; which may be computationally prohibitive in a real-time application. Therefore, in practice, one must often resort to simpler suboptimal procedures. One such simplification is to formulate the 2-D problem as two 1-D problems, i.e., to consider a *decoupled* data model. This is the topic of Chapter 4. A natural first question to ask is: what is the price in estimation accuracy of using the simpler 1-D/1-D data model? The CRB on the DOA and Doppler frequency estimation errors are derived for the decoupled data model, and shown to agree asymptotically with the corresponding bound for the full 2-D model. Furthermore, these asymptotic bounds for the multi-source case are shown to agree with the bounds for the single-source case. Consequently, a 1-D/1-D estimator that achieves the 2-D CRB in large samples is presented. Issues related to the implementation of the WLS estimators are also discussed. Finally, the first part of the thesis ends with some numerical examples and simulation results to corroborate the theoretical analysis.

2.2 Literature

An edited version of the material presented in the first part of this thesis will also be published as a book chapter in [AVE]. Parts of the material have previously been published in [Ath99, Ath01a]. The results presented herein are extensions of related work that has been reported previously by other authors. The re-formulation of the 2-D DOA/Doppler estimation problem as two sequential 1-D problems was first presented in [SS98]. Consistent and asymptotically efficient 1-D and 2-D MLEs with corresponding CRBs were derived. It was found that the 1-D and 2-D estimators have the same asymptotic performance. These results rely on assumptions of temporally white and spatially colored noise. However, they considered only the case of a single source. In this thesis we will extend some of the results in [SS98] to the case of multiple sources and also allow for temporally colored noise. More specifically, the space-time noise covariance is assumed to have a Kronecker structure. Other approaches to simplifying a 2-D frequency estimation problem into 1-D problems have been reported in e.g. [Hua92, RZZ93, SSM93, ZLS97].

In [ZF96] a related discussion about CRBs for spatio-temporal versus spatial-only processing was presented. Therein, it was shown that the CRB for a multi-source spatio-temporal model equals the CRB for a single-source

spatial model if the "source matrices" (covariance matrix of the source signals and the derivatives) are diagonal. These results are based on assumptions of spatio-temporally white noise. In this thesis we extend this result to colored noise, and show that it holds asymptotically when the source signals are complex sinusoids.

Additional results regarding the 2-D CRB and MLE include [War95], where the 2-D DOA/Doppler CRB and MLE were derived for the case of a single target in arbitrarily colored noise. CRB, least-squares (LS) and ML estimation in the context of 2-D frequency estimation have been treated in [Hua92, Cla93, RZZ94, LSZ96, KM96]; where the case of multiple signals in spatio-temporally white noise was considered. However, the finite sample multi-source CRBs are complicated and do not offer much insight. More easily interpretable, asymptotic results were given in [Cla93], where the asymptotic 2-D CRB was derived, as the amount of data tends to infinity in *both* frequency dimensions. Uniform sampling was assumed in both dimensions. However, in some applications, it is reasonable to expect that the amount of data is large in only one dimension. For instance, in a radar system employing an array antenna, it is usually more costly to use many antenna elements than to use many pulses. In addition, the sampling is, in many applications, not uniform. Nonuniform sampling is a means of combatting ambiguities that occur when the sampling is sparse. In this thesis we generalize the results of [Cla93] to nonuniform sampling and colored noise. Furthermore, we relax the condition that the amount of data is large in both dimensions and show that it is sufficient with one dimension, provided the processing is performed in the right order.

In [LSZ96, DN02] it was shown that the frequency dimensions are uncoupled also for finite samples. However, they used a more general signal model in which the matrix of signal amplitudes is full. For the problem considered in this thesis, the matrix of signal amplitudes is diagonal. This mean that to each DOA there corresponds only one Doppler frequency, and vice versa. A full matrix implies that each DOA corresponds to all Doppler frequencies and vice versa, see further the next subsection. In such a mixed-up scenario there is apparently nothing to gain using a 2-D model over 1-D models. This agrees with the observations in [ZF96] and [LR93] that nothing can be gained by using spatio-temporal processing if the signals are coherent. Herein we show that for a signal model with a diagonal signal amplitude matrix, the performance gain when using the full 2-D model can be substantial in finite samples.

We would also like to note an interesting relationship with the DOA CRB for signals with known waveforms derived in [LR93, LHSV95]. Therein it was shown that the multi-source CRB equals the single-source CRB if

the signals are uncorrelated. For the problem treated here, the signals are neither known nor uncorrelated. However, when the amount of temporal data is large the signals are almost uncorrelated and the waveforms estimates are very accurate. Therefore, as will be shown later on, the multi-source CRB will reach the single-source bound asymptotically also in our problem.

Recently, there has been an increased interest in asymptotic analysis of the 2-D frequency estimation problem. A number of publications have appeared. Here, we will only mention the most relevant to the work presented in this thesis. In [MS02], the asymptotic 2-D CRB is derived for the case of stationary colored noise. This implies a Toeplitz-block-Toeplitz structure of the space-time covariance matrix. Further assumptions in [MS02] are that the sampling is uniform and that the amount of data tends to infinity in both frequency dimensions. In [KN01, Eri02, CF02, MS02] the asymptotic distributions of 2-D nonlinear LS estimators in colored noise are derived. These analyses are also valid for non-Gaussian noise. For Gaussian noise, the estimators are shown to be asymptotically efficient.

To summarize, the main contribution of Part I in this thesis is an asymptotic analysis which shows that it is possible to obtain 2-D efficient estimates in both frequency dimensions when the amount of data is large in only one dimension. For this to hold, the processing must be performed in the right order, i.e. the "asymptotic dimension" should be processed first. Furthermore, we show that not only the frequency dimensions decouple, but also the different sources. This means that, under the stated assumptions, the single-source 2-D CRB can be attained in a multi-source scenario using two 1-D estimators. Moreover, the 1-D M-source estimator in the second stage reduces to M single-source estimators, where M is the number of sources.

2.3 Problem Formulation

In this section we present the data model, set the notation, and state the assumptions that our results are based upon. Consider a pulsed Doppler radar system employing an array antenna with K sensors. Within a coherent processing interval the radar transmits a series of N pulses. During this time the RCSs of the targets in the field of view are assumed to be constant. The pulse-to-pulse return baseband signals are modeled as complex sinusoids. It is assumed that the targets are in the far-field of the array and that the received signal waveforms are narrowband relative to the carrier frequency and array size. The array output for a particular range bin, in which M targets are present, can then be modeled by the $K \times 1$ vector

$$\mathbf{y}(n) = \mathbf{A}(\boldsymbol{\theta})\mathbf{B}\mathbf{s}(\boldsymbol{\omega}, n) + \mathbf{n}(n), \qquad n = 1, \ldots, N, \tag{2.1}$$

where

$$\begin{aligned} \mathbf{A}(\boldsymbol{\theta}) &= \left[\begin{array}{ccc} \mathbf{a}(\theta_1) & \ldots & \mathbf{a}(\theta_M) \end{array} \right] && (2.2) \\ \mathbf{B} &= \mathrm{diag}\left[\begin{array}{ccc} b_1 & \ldots & b_M \end{array} \right] && (2.3) \\ \mathbf{s}(\boldsymbol{\omega}, n) &= \left[\begin{array}{ccc} e^{j\omega_1 t_n} & \ldots & e^{j\omega_M t_n} \end{array} \right]^T. && (2.4) \end{aligned}$$

Here, $\mathbf{a}(\theta_m)$ is the complex $K \times 1$ steering vector that models the array response to a unit waveform from the DOA θ_m, which is defined to be the angle w.r.t. boresight of the array. The notation $\mathrm{diag}[\ b_1 \ \ldots \ b_M\ ]$ means a diagonal matrix with b_m as the m-th diagonal element. Furthermore, b_m denotes the complex amplitude and ω_m the Doppler frequency of the m-th source. The complex $K \times 1$ vector $\mathbf{n}(n)$ accounts for receiver noise as well as modelling errors and unmodeled signals such as jammers and clutter. Parameter estimates will be denoted $\hat{\theta}_m, \hat{\omega}_m$, and $\hat{b}_m$ respectively. The model in (2.1) can be put in matrix form according to

$$\mathbf{Y} = \mathbf{A}(\boldsymbol{\theta})\mathbf{B}\mathbf{S}(\boldsymbol{\omega}) + \mathbf{N}, \tag{2.5}$$

where

$$\begin{aligned} \mathbf{Y} &= \left[\begin{array}{ccc} \mathbf{y}(1) & \ldots & \mathbf{y}(N) \end{array} \right] && (2.6) \\ \mathbf{S}(\boldsymbol{\omega}) &= \left[\begin{array}{ccc} \mathbf{s}(\boldsymbol{\omega}, 1) & \ldots & \mathbf{s}(\boldsymbol{\omega}, N) \end{array} \right] && (2.7) \\ \mathbf{N} &= \left[\begin{array}{ccc} \mathbf{n}(1) & \ldots & \mathbf{n}(N) \end{array} \right]. && (2.8) \end{aligned}$$

The chief problem of interest herein is the estimation of the DOAs and Doppler frequencies given the noisy observations $\mathbf{Y}$. The complex amplitudes and the noise parameters are also unknown. However, these parameters are decoupled from the estimation of the DOAs and Doppler frequencies, and since they are usually not of great interest in radar systems, we treat them as nuisance parameters. The number of signals is assumed to be known, or estimated by some method. Estimation of the number of signals has been treated in e.g. [WK85, VOK91, Eri02]. In the 2-D problem, it is assumed that the $\{\omega_m, \theta_m\}$ pairs for different targets are distinct, while a somewhat more restrictive assumption is required in the 1/D-1/D problem, see Section 4.3. The noise is assumed to be multivariate complex Gaussian distributed with zero mean and second order moments

$$\begin{aligned} \mathrm{E}[\mathbf{n}(n)\mathbf{n}^H(l)] &= c_{nl}\mathbf{Q} && (2.9) \\ \mathrm{E}[\mathbf{n}(n)\mathbf{n}^T(l)] &= \mathbf{0} \quad \forall\, n, l. && (2.10) \end{aligned}$$

The $K \times K$ matrix $\mathbf{Q}$ models the spatial correlation of the noise, whereas the temporal correlation is included in the $N \times N$ matrix $\mathbf{C}$:

$$\mathbf{C} = \begin{bmatrix} c_{11} & \dots & c_{1N} \\ \vdots & & \vdots \\ c_{N1} & \dots & c_{NN} \end{bmatrix} \tag{2.11}$$

The structure implied by (2.9) means that the covariance matrix of the vectorized noise matrix takes the form

$$\mathrm{E}[\,\mathrm{vec}(\mathbf{N})\,\mathrm{vec}^H(\mathbf{N})] = \mathbf{C} \otimes \mathbf{Q}\,, \tag{2.12}$$

where $\mathrm{vec}(\mathbf{N})$ is formed by stacking the columns of $\mathbf{N}$ and $\otimes$ denotes the Kronecker matrix product [Bre78]. Now, $\mathbf{Y}$ contains $2NK$ real-valued observations, whereas $\mathbf{C}$ and $\mathbf{Q}$ together contain $N^2 + K^2$ real parameters (assuming only the Hermitian property is exploited). It is therefore not possible to estimate both $\mathbf{C}$ and $\mathbf{Q}$ from data. If more structure is imposed, for example that $\mathbf{C}$ is Toeplitz due to the noise being stationary, the number of free parameters reduces. We will not discuss this any further here, but simply assume a uniquely identifiable parameterization. More precisely, if $N > K$, $\mathbf{C}$ must be known, whereas $\mathbf{Q}$ is assumed known if $K > N$. Knowledge of $\mathbf{C}$ or $\mathbf{Q}$ is normally not available a priori, but estimates may be obtained from so called *secondary data*, i.e. data which are free from target signals and only contain noise and interference with the same statistics as the primary data set. In pulsed Doppler radar, secondary data may be obtained from range bins adjacent to the primary bin under consideration.

The noise parameterization or the possible knowledge of $\mathbf{C}$ or $\mathbf{Q}$ does not affect the CRB on the signal parameters as long as the noise parameterization is independent of the signal parameters and the model is identifiable. This is due to the fact that the Fisher information matrix (FIM) then becomes block-diagonal w.r.t. the signal and noise parameters.

The imposed Kronecker structure on the space-time covariance matrix may not be realistic in all applications. Consider a case that is realistic for an airborne radar system. The noise term is assumed to be composed of three terms; thermal noise, jamming, and clutter. A reasonable assumption on the structure of the thermal noise covariance is that it is spatio-temporally white, while the jamming covariance can be considered temporally white but spatially colored. In [War94] it was shown that the clutter space-time covariance matrix can be modeled as a sum of the contributions of different clutter patches where each term in the sum has a Kronecker structure. However, this does not imply that the sum has a Kronecker structure. Therefore, in

this case, the noise term will not have a space-time covariance matrix with a Kronecker structure

A motivation for assuming a Kronecker structure could instead be, besides computational simplicity, that it accommodates spatially *or* temporally colored noise in the same model. As alluded to above, the covariance matrix cannot be estimated anyway if both $\mathbf{C}$ and $\mathbf{Q}$ are unknown, and no other structure is imposed. Therefore, we will in the sequel assume that either $\mathbf{C}$ or $\mathbf{Q}$ is known. The dimension in which the covariance is known is then reduced to the white noise case by pre-whitening.

It is further assumed that the array manifold is unambiguous, i.e. that steering vectors $\mathbf{a}(\theta_m)$ corresponding to distinct DOAs $\theta_m,\ m = 1, \ldots, M$ are linearly independent. This is necessary for the DOAs to be uniquely identifiable from the data (actually, if the ω_ms are all distinct, one can relax this assumption). For the asymptotic properties of estimators to apply, it is required that $\mathbf{a}(\theta)$ possesses bounded derivatives up to third order. The first-order derivative is denoted as

$$\mathbf{d}_\theta(\theta_m) = \left.\frac{\partial \mathbf{a}(\theta)}{\partial \theta}\right|_{\theta=\theta_m}, \tag{2.13}$$

and the matrix of steering vector derivatives is

$$\mathbf{D}_\theta = \left[\mathbf{d}_\theta(\theta_1), \ldots, \mathbf{d}_\theta(\theta_M)\right]. \tag{2.14}$$

When deriving the CRB expressions, it is convenient to introduce a row-wise slicing of the signal matrix:

$$\mathbf{S}_N(\boldsymbol{\omega}) = \mathbf{S}^T(\boldsymbol{\omega}) = \left[\mathbf{s}_N(\omega_1), \ldots, \mathbf{s}_N(\omega_M)\right] \tag{2.15}$$

$$\mathbf{s}_N(\omega_m) = \left[s(\omega_m, 1), \ldots, s(\omega_m, N)\right]^T. \tag{2.16}$$

The corresponding derivatives are denoted as

$$\mathbf{d}_\omega(\omega_m) = \left.\frac{\partial \mathbf{s}_N(\omega)}{\partial \omega}\right|_{\omega=\omega_m} \tag{2.17}$$

$$\mathbf{D}_\omega = \left[\mathbf{d}_\omega(\omega_1), \ldots, \mathbf{d}_\omega(\omega_M)\right]. \tag{2.18}$$

The argument of the various vectors and matrices will frequently be dropped, when there is no risk of confusion.

Later on, the asymptotic estimation performance as $K \to \infty$ and $N \to \infty$ will be investigated. The results for the $K \to \infty$ case rely on the assumption that the following matrices

$$\lim_{K\to\infty} \frac{1}{K}\mathbf{A}^H\mathbf{A}, \quad \lim_{K\to\infty} \frac{1}{K^2}\mathbf{A}^H\mathbf{D}_\theta, \quad \lim_{K\to\infty} \frac{1}{K^3}\mathbf{D}_\theta^H\mathbf{D}_\theta \tag{2.19}$$

exist and are diagonal. A similar assumption appeared in [Vib95]. A natural question to ask is, for which class of array configurations are these assumptions satisfied? This question is far from trivial. It is easy to verify, however, that they are satisfied for ULAs. We will not pursue this topic further, but adopt (2.19) when dealing with the case of large K. A similar assumption involving $\mathbf{S}_N$ and $\mathbf{D}_\omega$ is necessary in the case of large N, see Section 3.3. The latter is valid when the temporal sampling is uniform.

A comment on the validity of the data model is in order. Letting $K \to \infty$ may cause the data model to become invalid since the assumptions about far-field and narrowband emitters are violated. Also, the array gain will tend to infinity. It should be understood that the asymptotic analysis is merely a mathematical tool to obtain tractable solutions. We will find by simulation that the asymptotic results are valid with good approximation also for reasonably sized arrays, for which the model is not unrealistic.

We would like to end this introduction by explicitly showing the close connection to 2-D frequency estimation. First assume that a ULA with omnidirectional elements that have unity gain is employed. The steering vector can then according to (1.14) be expressed as

$$\mathbf{a}(\theta) = \left[\begin{array}{cccc} e^{j\phi_1} & e^{j\phi_2} & \dots & e^{j\phi_K} \end{array} \right]^T, \tag{2.20}$$

where $\phi_k \triangleq -2\pi(k-1)d\sin(\theta)/\lambda$. If we further assume uniform temporal sampling, i.e. $t_n = n, n = 1, \dots, N$, the received signal vector can according to (2.1) be written as

$$\mathbf{y}(n) \;=\; \sum_{m=1}^{M} b_m \mathbf{a}(\theta_m) e^{j\omega_m n} + \mathbf{n}(n), \quad n = 1, \dots, N. \tag{2.21}$$

The k-th element in $\mathbf{y}(n)$ is then given by

$$y_k(n) \;=\; \sum_{m=1}^{M} b_m e^{j(\phi_m k + \omega_m n)} + n_k(n), \quad k = 1, \dots, K,\ n = 1, \dots, N. \tag{2.22}$$

In this model, $\{\phi_m, \omega_m\}$ are 2-D frequencies. This model is often used in 2-D frequency estimation, see e.g. [Hua92, LSZ96, MS02]. Another model that occurs is, see e.g. [RZZ94, LSZ96],

$$y_k(n) \;=\; \sum_{\ell=1}^{L} \sum_{m=1}^{M} b_{\ell m} e^{j(\phi_\ell k + \omega_m n)} + n_k(n), \quad k = 1, \dots, K,\ n = 1, \dots, N. \tag{2.23}$$

For the case of $L = M$, this model corresponds to the model in (2.1) with a full $\mathbf{B}$ matrix. We deal only with the case of diagonal $\mathbf{B}$ in the thesis. However, the model employed herein is more general than the one in (2.22), since we allow nonuniform sampling and a more general form of the steering vector according to

$$\mathbf{a}(\theta) = \left[\begin{array}{ccc} H_1(\theta)\, e^{j\, 2\pi d_1 \sin(\theta)/\lambda} & \ldots & H_K(\theta)\, e^{j\, 2\pi d_K \sin(\theta)/\lambda} \end{array} \right]^T, \qquad (2.24)$$

where $d_k, k = 1, \ldots, K$, are the element positions.

Chapter 3

Estimation Using the 2-D Model

In this chapter we will present performance bounds and methods assuming the full 2-D data model is exploited. This is clearly advantageous from a performance point of view, but leads to computationally expensive estimation methods. Therefore, the next chapter is devoted to methods based on decoupled 1-D data models.

3.1 General CRB for Parameterized Signals

In order to have a benchmark, the presentation begins by studying the achievable performance. We first derive the CRB for a general nonlinear signal parameterization, and later specialize to the case of DOA/Doppler estimation which is of major concern herein. Therefore, let each of the M source signals be uniquely described by a known, parameter dependent, complex function $x(\boldsymbol{\alpha}_m, n)\,,\ m = 1, \ldots, M$, where $\boldsymbol{\alpha}_m$ is an r-dimensional column vector of signal parameters for the m-th source, functionally independent of the DOAs. The number of signal parameters per source, r, is finite. Note that this corresponds to a deterministic, parameter dependent signal model. Examples are the sinusoidal signal model and polynomial phase signals.

Now, the K-vector, $\mathbf{y}(n)$, of sensor outputs, observed in additive noise, $\mathbf{n}(t)$, is modeled by the following generalization of (2.1)

$$\mathbf{y}(n) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(\boldsymbol{\alpha}, n) + \mathbf{n}(n)\,,\ n = 1, \ldots, N\,, \tag{3.1}$$

where

$$\mathbf{x}(\boldsymbol{\alpha}, n) = [x(\boldsymbol{\alpha}_1, n), \ldots, x(\boldsymbol{\alpha}_M, n)]^T\,. \tag{3.2}$$

Collect the observations in the KN-vector $\mathbf{y}_N$

$$\mathbf{y}_N = \text{ vec}(\mathbf{Y}) = \left[\mathbf{y}^T(1), \ldots, \mathbf{y}^T(N)\right]^T\,. \tag{3.3}$$

Since the signal is deterministic and the noise is complex Gaussian with covariance matrix given by (2.12), $\mathbf{y}_N$ is distributed as

$$\mathbf{y}_N \in \mathcal{N}(\mathbf{m}(\boldsymbol{\theta}, \boldsymbol{\alpha}), \mathbf{C} \otimes \mathbf{Q}) \,, \tag{3.4}$$

where the mean vector is given by

$$\mathbf{m}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \left[(\mathbf{A}(\boldsymbol{\theta})\mathbf{x}(\boldsymbol{\alpha}, 1))^T, \ldots, (\mathbf{A}(\boldsymbol{\theta})\mathbf{x}(\boldsymbol{\alpha}, N))^T \right]^T . \tag{3.5}$$

The likelihood function is therefore given by (see e.g. [Kay93, Section 15.7])

$$p(\mathbf{y}_N; \boldsymbol{\eta}) = \frac{1}{\pi^{KN} \left| \mathbf{C} \otimes \mathbf{Q} \right|} \exp\left\{ -(\mathbf{y}_N - \mathbf{m})^H (\mathbf{C} \otimes \mathbf{Q})^{-1} (\mathbf{y}_N - \mathbf{m}) \right\} , \tag{3.6}$$

where $\boldsymbol{\eta}$ is a real vector representing the unknown parameters. Since the noise covariance matrix does not depend on the signal parameters, the FIM has a block diagonal structure. This implies that the CRB for the signal parameters is the same whether or not the noise covariance is known (of course, all unknown parameters must be uniquely identifiable). Estimation of the noise properties will therefore be ignored in this section, and only the FIM and CRB for the signal parameters will be derived. Thus, let the unknown parameter vector be

$$\boldsymbol{\eta} \;=\; \left[\; \alpha_{1,1} \;\; \ldots \;\; \alpha_{M,1} \;\; \ldots \;\; \alpha_{1,r} \;\; \ldots \;\; \alpha_{M,r} \;\; \theta_1 \;\; \ldots \;\; \theta_M \; \right]^T ,$$

where $\alpha_{k,l}$ is the l-th parameter in $\boldsymbol{\alpha}_k$. The following general expression for the (r, s)-th element of the FIM is obtained from [Kay93, Section 15.7]

$$J_{rs} = 2\mathrm{Re}\left\{ \frac{\partial \mathbf{m}^H(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \eta_r} \; (\mathbf{C} \otimes \mathbf{Q})^{-1} \; \frac{\partial \mathbf{m}(\boldsymbol{\theta}, \boldsymbol{\alpha})}{\partial \eta_s} \right\} . \tag{3.7}$$

It is now straightforward to derive the following.

Theorem 3.1 *The CRB for* $\boldsymbol{\eta}$ *is*

$$\mathrm{E}[(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})^T] \geq \mathbf{J}^{-1} \tag{3.8}$$

where $\hat{\boldsymbol{\eta}}$ *represents any unbiased estimator of* $\boldsymbol{\eta}$*, and the notation* $\mathbf{A} \geq \mathbf{B}$ *means that* $\mathbf{A} - \mathbf{B}$ *is positive semidefinite. The FIM has the block structure*

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{1,1} & \ldots & \mathbf{J}_{1,r} & \mathbf{J}_{1,r+1} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{J}_{1,r}^T & \ldots & \mathbf{J}_{r,r} & \mathbf{J}_{r,r+1} \\ \mathbf{J}_{1,r+1}^T & \ldots & \mathbf{J}_{r,r+1}^T & \mathbf{J}_{r+1,r+1} \end{bmatrix} ,$$

where the blocks are given by

$$\begin{aligned}
\mathbf{J}_{k,l} &= 2\mathrm{Re}\left[({\mathbf{X}'_k}^H \mathbf{C}^{-1} \mathbf{X}'_l) \odot (\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A}) \right] \ , k,l = 1 \dots r \, , \\
\mathbf{J}_{k,r+1} &= 2\mathrm{Re}\left[({\mathbf{X}'_k}^H \mathbf{C}^{-1} \mathbf{X}) \odot (\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{D}_\theta) \right] \ , k = 1 \dots r \, , \\
\mathbf{J}_{r+1,r+1} &= 2\mathrm{Re}\left[(\mathbf{X}^H \mathbf{C}^{-1} \mathbf{X}) \odot (\mathbf{D}_\theta^H \mathbf{Q}^{-1} \mathbf{D}_\theta) \right] .
\end{aligned}$$

Here, $\odot$ denotes element-wise matrix multiplication (Hadamard product), and, further, the following definitions are utilized

$$\begin{aligned}
\mathbf{X} &= \left[\begin{array}{ccc} \mathbf{x}_1 & \dots & \mathbf{x}_M \end{array} \right] \\
\mathbf{X}'_k &= \left[\begin{array}{ccc} \dfrac{\partial \mathbf{x}_1}{\partial \alpha_{1,k}} & \dots & \dfrac{\partial \mathbf{x}_M}{\partial \alpha_{M,k}} \end{array} \right] ,
\end{aligned}$$

and

$$\mathbf{x}_k = \left[\begin{array}{cccc} x(\boldsymbol{\alpha}_k, 1) & \dots & x(\boldsymbol{\alpha}_k, N) \end{array} \right]^T .$$

The matrix $\mathbf{A}$ is the steering matrix corresponding to the true directions of arrival, and $\mathbf{D}_\theta$ the matrix of steering vector derivatives (2.14).

Proof: Straightforward calculations using (3.7). The details for the case $\mathbf{C} = \mathbf{I}$ are given in [Eri02].

3.2 CRB for the 2-D Model

The above result applies to a general parameterization of the emitter signals. We will now return to the application of main concern herein. Thus, let the signal be given as $\mathbf{x}(\boldsymbol{\alpha}, n) = \mathbf{B}\mathbf{s}(\boldsymbol{\omega}, n)$, i.e., the data model according to (2.1) is

$$\mathbf{y}(n) = \mathbf{A}(\boldsymbol{\theta})\mathbf{B}\mathbf{s}(\boldsymbol{\omega}, n) + \mathbf{n}(n) \, , \ n = 1, \dots, N. \tag{3.9}$$

Since we are mainly concerned with estimation of the DOA and Doppler parameters, only the bound for these estimates will be presented. The corresponding result for the complex amplitudes can be deduced from the proof.

Theorem 3.2 *Let $\hat{\boldsymbol{\omega}}$ and $\hat{\boldsymbol{\theta}}$ be unbiased estimates of the DOA and Doppler parameters. Then it holds that*

$$\mathrm{E} \left[\begin{array}{c} \hat{\boldsymbol{\omega}} - \boldsymbol{\omega} \\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \end{array} \right] \left[\begin{array}{c} \hat{\boldsymbol{\omega}} - \boldsymbol{\omega} \\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \end{array} \right]^T \geq \left[\begin{array}{cc} \mathbf{C}_{\omega\omega} & \mathbf{C}_{\omega\theta} \\ \mathbf{C}_{\omega\theta}^T & \mathbf{C}_{\theta\theta} \end{array} \right] . \tag{3.10}$$

The blocks in the above matrix are given by

$$\begin{aligned}
\mathbf{C}_{\omega\omega} &= \left(2\mathbf{R}_{\omega}^{T}\mathbf{P}_{\mathbf{R}_\theta}^{\perp}\mathbf{R}_{\omega}\right)^{-1} && (3.11)\\
\mathbf{C}_{\theta\theta} &= \left(2\mathbf{R}_{\theta}^{T}\mathbf{P}_{\mathbf{R}_\omega}^{\perp}\mathbf{R}_{\theta}\right)^{-1} && (3.12)\\
\mathbf{C}_{\omega\theta} &= -\left(2\mathbf{R}_{\omega}^{T}\mathbf{P}_{\mathbf{R}_\theta}^{\perp}\mathbf{R}_{\omega}\right)^{-1}\left(\mathbf{R}_{\theta}^{\dagger}\mathbf{R}_{\omega}\right)^{T}, && (3.13)
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{P}_{\mathbf{R}_\theta}^{\perp} &= \mathbf{I}-\mathbf{R}_{\theta}\mathbf{R}_{\theta}^{\dagger}, \quad \mathbf{R}_{\theta}^{\dagger}=(\mathbf{R}_{\theta}^{T}\mathbf{R}_{\theta})^{-1}\mathbf{R}_{\theta}^{T} && (3.14)\\
\mathbf{P}_{\mathbf{R}_\omega}^{\perp} &= \mathbf{I}-\mathbf{R}_{\omega}\mathbf{R}_{\omega}^{\dagger}, \quad \mathbf{R}_{\omega}^{\dagger}=(\mathbf{R}_{\omega}^{T}\mathbf{R}_{\omega})^{-1}\mathbf{R}_{\omega}^{T}. && (3.15)
\end{aligned}$$

and the following definitions have been utilized

$$\mathbf{R}_{\omega}=\begin{bmatrix}\mathrm{Re}[\mathbf{P}_{\Upsilon}^{\perp}\boldsymbol{\Upsilon}_{\omega}\mathbf{B}]\\ \mathrm{Im}[\mathbf{P}_{\Upsilon}^{\perp}\boldsymbol{\Upsilon}_{\omega}\mathbf{B}]\end{bmatrix}, \quad \mathbf{R}_{\theta}=\begin{bmatrix}\mathrm{Re}[\mathbf{P}_{\Upsilon}^{\perp}\boldsymbol{\Upsilon}_{\theta}\mathbf{B}]\\ \mathrm{Im}[\mathbf{P}_{\Upsilon}^{\perp}\boldsymbol{\Upsilon}_{\theta}\mathbf{B}]\end{bmatrix}, \qquad (3.16)$$

where

$$\begin{aligned}
\mathbf{P}_{\Upsilon}^{\perp} &= \mathbf{I}-\boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}^{H}\boldsymbol{\Upsilon})^{-1}\boldsymbol{\Upsilon}^{H} && (3.17)\\
\boldsymbol{\Upsilon} &= (\mathbf{C}^{-1/2}\mathbf{S}_N)\diamond(\mathbf{Q}^{-1/2}\mathbf{A}) && (3.18)\\
\boldsymbol{\Upsilon}_{\omega} &= (\mathbf{C}^{-1/2}\mathbf{D}_{\omega})\diamond(\mathbf{Q}^{-1/2}\mathbf{A}) && (3.19)\\
\boldsymbol{\Upsilon}_{\theta} &= (\mathbf{C}^{-1/2}\mathbf{S}_N)\diamond(\mathbf{Q}^{-1/2}\mathbf{D}_{\theta}). && (3.20)
\end{aligned}$$

Here, $\diamond$ denotes the Khatri-Rao product (column-wise Kronecker product, see [Bre78]) and $\mathbf{M}^{1/2}$ denotes a Hermitian square-root factor of the matrix $\mathbf{M}$, i.e. $\mathbf{M}=\mathbf{M}^{1/2}\mathbf{M}^{1/2}$ and $\mathbf{M}^{-1}=\mathbf{M}^{-1/2}\mathbf{M}^{-1/2}$. Furthermore, the signal matrix $\mathbf{S}_N=\mathbf{S}^T$ and the corresponding matrix of derivatives $\mathbf{D}_{\omega}$ are defined in (2.15)–(2.18).

Proof: See Appendix 3A.

3.3 Asymptotic CRB for the 2-D Model

Although explicit expressions for the bounds on the DOA and Doppler estimation errors are given above, they might appear somewhat unwieldy. As we will see, much simpler results are obtained in the large sample case, assuming either K or N to be large.

Large Number of Antenna Elements

The first case considered assumes that K is "large enough". The following additional assumptions are necessary to simplify the bound:

- The matrices $\lim_{K\to\infty}\frac{1}{K}\mathbf{A}^H\mathbf{A}$, $\lim_{K\to\infty}\frac{1}{K^2}\mathbf{A}^H\mathbf{D}_\theta$, $\lim_{K\to\infty}\frac{1}{K^3}\mathbf{D}_\theta^H\mathbf{D}_\theta$ exist and are all diagonal. This is valid, for example, for a ULA.
- The noise is spatially white and (possibly) temporally colored, i.e. $\mathbf{Q}=\mathbf{I}$.

We then have the following result:

Theorem 3.3 *For $K\to\infty$, the 2-D DOA and Doppler CRB matrices $\mathbf{C}_{\theta\theta}$ and $\mathbf{C}_{\omega\omega}$ are both diagonal and $\mathbf{C}_{\omega\theta}=\mathbf{0}$. The m-th diagonal elements are obtained by evaluating the expressions*

$$c_\theta = \frac{1}{2|b|^2\ \mathbf{s}_N^H\mathbf{C}^{-1}\mathbf{s}_N\ \mathbf{d}_\theta^H\mathbf{P}_{\mathbf{a}}^\perp\mathbf{d}_\theta} \tag{3.21}$$

$$c_\omega = \frac{1}{2|b|^2\ \mathbf{d}_\omega^H\mathbf{C}^{-1/2}\mathbf{P}^\perp_{\mathbf{C}^{-1/2}\mathbf{s}_N}\mathbf{C}^{-1/2}\mathbf{d}_\omega\ \mathbf{a}^H\mathbf{a}} \tag{3.22}$$

at the true parameter values ω_m,θ_m and b_m. Here, the projection matrices are defined as $\mathbf{P}_{\mathbf{a}}^\perp=\mathbf{I}-\frac{\mathbf{a}\mathbf{a}^H}{\mathbf{a}^H\mathbf{a}}$ and $\mathbf{P}^\perp_{\mathbf{C}^{-1/2}\mathbf{s}_N}=\mathbf{I}-\frac{\mathbf{C}^{-1/2}\mathbf{s}_N\mathbf{s}_N^H\mathbf{C}^{-1/2}}{\mathbf{s}_N^H\mathbf{C}^{-1}\mathbf{s}_N}$. The asymptotic bounds for $\hat\theta_m$ and $\hat\omega_m$ are equal to the corresponding expressions for the single-source case.

Proof: See Appendix 3B.

As one could expect, the asymptotic bound for $\hat\theta_m$ is equal to the single-source bound. This is because the array aperture increases without limit as $K\to\infty$. What is perhaps more surprising is that the bound on $\hat\omega_m$ also reduces to the single-source case, although no assumption on N is made. Apparently, it is sufficient to resolve the sources in one dimension to also decouple the problem in the other dimension. As we will see later, the asymptotic 2-D CRB expressions are in fact attainable also using the decoupled 1-D/1-D signal model.

Large Number of Pulses

We now turn to the dual case $N\to\infty$. The extra assumptions above are now replaced by:

- The matrices $\lim_{N\to\infty} \frac{1}{N}\mathbf{S}_N^H\mathbf{S}_N$, $\lim_{N\to\infty} \frac{1}{N^2}\mathbf{D}_\omega^H\mathbf{S}_N$, $\lim_{N\to\infty} \frac{1}{N^3}\mathbf{D}_\omega^H\mathbf{D}_\omega$ exist and are all diagonal. This is true for uniformly sampled data.

- The noise is (possibly) spatially colored but temporally white, i.e. $\mathbf{C} = \mathbf{I}$.

Theorem 3.4 *For $N \to \infty$, the 2-D DOA and Doppler CRB matrices $\mathbf{C}_{\theta\theta}$ and $\mathbf{C}_{\omega\omega}$ are both diagonal and $\mathbf{C}_{\omega\theta} = \mathbf{0}$. The m-th diagonal elements are obtained by evaluating the expressions*

$$c_\theta = \frac{1}{2|b|^2\, \mathbf{s}_N^H\mathbf{s}_N\, \mathbf{d}_\theta^H\mathbf{Q}^{-1/2}\mathbf{P}^\perp_{\mathbf{Q}^{-1/2}\mathbf{a}}\mathbf{Q}^{-1/2}\mathbf{d}_\theta} \tag{3.23}$$

$$c_\omega = \frac{1}{2|b|^2\mathbf{d}_\omega^H\mathbf{P}^\perp_{\mathbf{s}_N}\mathbf{d}_\omega\, \mathbf{a}^H\mathbf{Q}^{-1}\mathbf{a}} \tag{3.24}$$

at the true parameter values ω_m, θ_m and b_m. The asymptotic bounds for $\hat{\theta}_m$ and $\hat{\omega}_m$ are equal to the corresponding expressions for the single target case.

Proof: See Appendix 3B.

Also in this case, the asymptotic CRB reduces to the single target case. It is interesting to note that the asymptotic CRB expressions for the case of large K and large N coincide when the noise is both spatially and temporally white.

Simpler expressions for the asymptotic 2-D CRB can be obtained if we consider the special case of a ULA with one-half wavelength element separation, uniform temporal sampling and spatio-temporally white noise. From the theorems above the following expressions are readily obtained

$$c_\theta = \frac{6\sigma^2}{NK\left(K^2-1\right)|b|^2} \cdot \frac{1}{\pi^2\cos^2\theta} \tag{3.25}$$

$$c_\omega = \frac{6\sigma^2}{NK\left(N^2-1\right)|b|^2}, \tag{3.26}$$

where σ^2 is the noise variance. In this form, it is easy to see how pertinent system design parameters affect the ultimate estimation accuracy. Note that these expressions are consistent with the early results in [Bre61, Ban71, RB74] on the CRB for (spatial or temporal) frequency estimation of a single signal.

3.4 2-D Maximum Likelihood Estimation

To apply the ML method we need the *likelihood function* of the observed data. The estimates are then obtained as the parameter values that maximize this function. An interpretation of this method is that the estimates are the parameter values that make the observed data most probable. The likelihood function for a general signal model is given in (3.6)

$$p(\mathbf{y}_N;\boldsymbol{\eta}) = \frac{1}{\pi^{KN}\,|\mathbf{C}\otimes\mathbf{Q}|}\exp\left\{-(\mathbf{y}_N-\mathbf{m})^H(\mathbf{C}\otimes\mathbf{Q})^{-1}(\mathbf{y}_N-\mathbf{m})\right\}\,, \quad (3.27)$$

where, here, $\mathbf{m} = \text{vec}(\mathbf{ABS})$. As alluded to previously, it is not possible to estimate both $\mathbf{C}$ and $\mathbf{Q}$ jointly with the signal parameters. We will therefore first consider the case of known noise color, and then the case where $N > K$ and only $\mathbf{C}$ is known. The analogous case where $K > N$ and $\mathbf{Q}$ is known can be obtained in a similar manner.

Known noise color

Let us first assume that $\mathbf{C}$ and $\mathbf{Q}$ are known (possibly up to scaling). One can then pre-whiten the data by the transformation

$$\mathbf{y}_{NW} = (\mathbf{C}\otimes\mathbf{Q})^{-1/2}\mathbf{y}_N = (\mathbf{C}^{-1/2}\otimes\mathbf{Q}^{-1/2})\mathbf{y}_N\,. \quad (3.28)$$

Using the properties of Kronecker products [Bre78], it is easy to see that (3.28) can efficiently be implemented as

$$\mathbf{y}_{NW} = \text{vec}\left(\mathbf{Q}^{-1/2}\mathbf{Y}\mathbf{C}^{-T/2}\right)\,. \quad (3.29)$$

After the pre-whitening operation, the noise becomes temporally and spatially white. To keep the notation simple, we will therefore take $\mathbf{C} = \sigma^2\mathbf{I}$ and $\mathbf{Q} = \mathbf{I}$, and keep in mind that the signal model needs to be modified when the noise is colored (see further Section 3.5). The unknown parameters are then $\boldsymbol{\theta}$, $\boldsymbol{\omega}$, σ^2, and $\mathbf{b} \triangleq \text{Diag}\{\mathbf{B}\}$. Maximizing the likelihood function is equivalent to minimizing the negative log-likelihood function, which in this case reduces to

$$\begin{aligned}
V_{ML}(\boldsymbol{\omega},\boldsymbol{\theta},\mathbf{b},\sigma^2) &= -\log p(\mathbf{y}_N;\boldsymbol{\omega},\boldsymbol{\theta},\mathbf{b},\sigma^2)\\
&= KN\log(\pi\sigma^2) + \frac{1}{\sigma^2}\sum_{n=1}^{N}\left[\mathbf{y}(n)-\mathbf{A}(\boldsymbol{\theta})\mathbf{B}\mathbf{s}(\boldsymbol{\omega},n)\right]^H\left[\mathbf{y}(n)-\mathbf{A}(\boldsymbol{\theta})\mathbf{B}\mathbf{s}(\boldsymbol{\omega},n)\right]\\
&= KN\log(\pi\sigma^2) + \frac{1}{\sigma^2}\text{Tr}\left\{[\mathbf{Y}-\mathbf{A}(\boldsymbol{\theta})\mathbf{B}\mathbf{S}(\boldsymbol{\omega})]^H[\mathbf{Y}-\mathbf{A}(\boldsymbol{\theta})\mathbf{B}\mathbf{S}(\boldsymbol{\omega})]\right\}\\
&= KN\log(\pi\sigma^2) + \frac{1}{\sigma^2}\|\mathbf{Y}-\mathbf{A}(\boldsymbol{\theta})\mathbf{B}\mathbf{S}(\boldsymbol{\omega})\|_F^2, \quad (3.30)
\end{aligned}$$

where $\mathbf{Y}$ and $\mathbf{S}(\boldsymbol{\omega})$ are defined in (2.6)–(2.7) and $\|\cdot\|_F$ denotes the Frobenius matrix norm. For fixed $\boldsymbol{\omega}, \boldsymbol{\theta}, \mathbf{b}$; the minimum of V_{ML} w.r.t. σ^2 is easily found as

$$\hat{\sigma}^2 = \frac{1}{KN}\|\mathbf{Y} - \mathbf{A}(\boldsymbol{\theta})\mathbf{B}\mathbf{S}(\boldsymbol{\omega})\|_F^2. \tag{3.31}$$

This estimate can then be substituted back into (3.30) (see [GP73] for motivation of such an operation). Thus,

$$V_{ML}(\boldsymbol{\omega}, \boldsymbol{\theta}, \mathbf{b}) = KN \log\left(\frac{\pi}{KN}\|\mathbf{Y} - \mathbf{A}(\boldsymbol{\theta})\mathbf{B}\mathbf{S}(\boldsymbol{\omega})\|_F^2\right) + KN. \tag{3.32}$$

Minimizing this expression w.r.t. the unknown parameters is equivalent to minimizing

$$V_{ML}(\boldsymbol{\omega}, \boldsymbol{\theta}, \mathbf{b}) = \|\mathbf{Y} - \mathbf{A}(\boldsymbol{\theta})\mathbf{B}\mathbf{S}(\boldsymbol{\omega})\|_F^2, \tag{3.33}$$

where the notation $V_{ML}(\boldsymbol{\omega}, \boldsymbol{\theta}, \mathbf{b})$ has been retained for convenience. We note in passing that minimizing (3.33) yields the MLE of $\boldsymbol{\omega}, \boldsymbol{\theta}$ and $\mathbf{b}$ also if σ^2 is known. Using the formula [Bre78]

$$\text{vec}\left[\mathbf{A}(\boldsymbol{\theta})\mathbf{B}\mathbf{S}(\boldsymbol{\omega})\right] = \left[\mathbf{S}^T(\boldsymbol{\omega}) \diamond \mathbf{A}(\boldsymbol{\theta})\right] \text{Diag}\{\mathbf{B}\}, \tag{3.34}$$

we can write (3.33) as

$$\begin{aligned} V_{ML}(\boldsymbol{\omega}, \boldsymbol{\theta}, \mathbf{b}) &= \|\,\text{vec}\left[\mathbf{Y} - \mathbf{A}(\boldsymbol{\theta})\mathbf{B}\mathbf{S}(\boldsymbol{\omega})\right]\|_F^2 \\ &= \left\|\,\text{vec}(\mathbf{Y}) - \left[\mathbf{S}^T(\boldsymbol{\omega}) \diamond \mathbf{A}(\boldsymbol{\theta})\right] \text{Diag}\{\mathbf{B}\}\right\|_F^2 \\ &= \|\mathbf{y}_N - \boldsymbol{\Upsilon}(\boldsymbol{\omega}, \boldsymbol{\theta})\mathbf{b}\|_F^2, \end{aligned} \tag{3.35}$$

where

$$\boldsymbol{\Upsilon}(\boldsymbol{\omega}, \boldsymbol{\theta}) = \mathbf{S}^T(\boldsymbol{\omega}) \diamond \mathbf{A}(\boldsymbol{\theta}). \tag{3.36}$$

In this form, $V_{ML}(\boldsymbol{\omega}, \boldsymbol{\theta}, \mathbf{b})$ is readily minimized w.r.t. $\mathbf{b}$ for fixed $\boldsymbol{\omega}$ and $\boldsymbol{\theta}$, yielding

$$\hat{\mathbf{b}} = \left(\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\right)^{-1}\boldsymbol{\Upsilon}^H\mathbf{y}_N. \tag{3.37}$$

Substituting into (3.35), results in the final expression

$$V_{ML}(\boldsymbol{\omega}, \boldsymbol{\theta}) = \left\|\mathbf{P}_{\Upsilon}^{\perp}(\boldsymbol{\omega}, \boldsymbol{\theta})\mathbf{y}_N\right\|_F^2, \tag{3.38}$$

where

$$\mathbf{P}_{\Upsilon}^{\perp}(\boldsymbol{\omega}, \boldsymbol{\theta}) = \mathbf{I} - \boldsymbol{\Upsilon}\left(\boldsymbol{\Upsilon}^H\boldsymbol{\Upsilon}\right)^{-1}\boldsymbol{\Upsilon}^H. \tag{3.39}$$

The ML solution for DOA/Doppler estimation is thus given by the solution to the 2-D nonlinear LS problem

$$\left\{\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\theta}}\right\} = \arg\min_{\boldsymbol{\omega},\boldsymbol{\theta}} \left\|\mathbf{P}_{\Upsilon}^{\perp}(\boldsymbol{\omega},\boldsymbol{\theta})\mathbf{y}_N\right\|_F^2 . \tag{3.40}$$

Although this optimization problem is actually $2M$-dimensional, we call it 2-D since it is a joint problem in two frequency dimensions, namely the spatial (DOA) and temporal (Doppler) frequencies. This is in contrast to the 1-D problems (which are M-dimensional optimization problems), where the estimation is performed in one of the frequency dimensions.

Known temporal noise covariance

Let us now turn to the more difficult case of colored noise, where $\mathbf{C}$ is known but $\mathbf{Q}$ is to be estimated. From (3.6), the negative log-likelihood function is in this case obtained as

$$\begin{aligned} V_{ML}(\boldsymbol{\omega},\boldsymbol{\theta},\mathbf{b},\mathbf{Q}) &= N\log|\mathbf{Q}| \\ &\quad + \left[\mathbf{y}_N - \operatorname{vec}(\mathbf{ABS})\right]^H (\mathbf{C}\otimes\mathbf{Q})^{-1}\left[\mathbf{y}_N - \operatorname{vec}(\mathbf{ABS})\right] , \end{aligned} \tag{3.41}$$

where parameter-independent terms have been ignored, and we have used that [Bre78]

$$|\mathbf{C}\otimes\mathbf{Q}| = |\mathbf{C}|^K|\mathbf{Q}|^N. \tag{3.42}$$

Again, applying formulas from [Bre78] involving the vectorization operator and the Kronecker product, one finds that

$$V_{ML}(\boldsymbol{\omega},\boldsymbol{\theta},\mathbf{b},\mathbf{Q}) = N\log|\mathbf{Q}| + \operatorname{Tr}\left\{\mathbf{Q}^{-1}(\mathbf{Y}-\mathbf{ABS})\mathbf{C}^{-T}(\mathbf{Y}-\mathbf{ABS})^H\right\}. \tag{3.43}$$

Minimizing (3.43) w.r.t. $\mathbf{Q}$ (see e.g. [MKB79, HS92, SS98]) results in

$$\hat{\mathbf{Q}} = \frac{1}{N}(\mathbf{Y}-\mathbf{ABS})\mathbf{C}^{-T}(\mathbf{Y}-\mathbf{ABS})^H. \tag{3.44}$$

Substituting back into (3.43) we obtain the MLE in the form

$$\{\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{b}}\} = \arg\min_{\boldsymbol{\omega},\boldsymbol{\theta},\mathbf{b}} \left|(\mathbf{Y}-\mathbf{ABS})\mathbf{C}^{-T}(\mathbf{Y}-\mathbf{ABS})^H\right| . \tag{3.45}$$

Unfortunately, no explicit solution of (3.45) with respect to $\mathbf{b}$ is known, so in this case one must resort to a nonlinear optimization over $4M$ real parameters. Reducing this to $2M$ is the motivation for the next method.

3.5 2-D Weighted Least Squares

The weighted least squares (WLS) principle is a common suboptimal technique in statistical signal processing [Kay93]. In our case, the criterion can be motivated by the second term of (3.43), which is expressed as

$$\begin{aligned} V_{WLS}(\boldsymbol{\omega}, \boldsymbol{\theta}, \mathbf{b}) &= \mathrm{Tr}\{\mathbf{W}_Q(\mathbf{Y} - \mathbf{ABS})\mathbf{W}_C(\mathbf{Y} - \mathbf{ABS})^H\} \\ &= \|\mathbf{W}_Q^{1/2}(\mathbf{Y} - \mathbf{ABS})\mathbf{W}_C^{1/2}\|_F^2 . \end{aligned} \tag{3.46}$$

Here, $\mathbf{W}_Q$ and $\mathbf{W}_C$ are arbitrary positive definite weighting matrices. The WLS approach itself makes no assumptions about the distribution or correlation of the data, so (3.46) can simply be motivated as selecting the model that fits the observed data best, in a (possibly weighted) least squares sense. However, from (3.43) we know that ML and WLS are equivalent under the stated assumptions on the noise, provided the weighting matrices are chosen as $\mathbf{W}_Q = \mathbf{Q}^{-1}$ and $\mathbf{W}_C = \mathbf{C}^{-T}$. Clearly, this requires the noise properties to be known. However, an interesting fact is that the optimal weighting matrices can be replaced by consistent estimates without changing the asymptotic properties of the WLS estimates. In essence, this is because the estimation of the parameters of a deterministic signal is decoupled from estimating the noise properties, i.e. the corresponding FIM is block diagonal [Kay93]. This means that asymptotically efficient estimates can be obtained using estimated noise properties. As previously explained, $\mathbf{Q}$ and $\mathbf{C}$ cannot be simultaneously estimated, unless some further structure is imposed. Thus, following the case presented above, it will be assumed that $\mathbf{C}$ is known and $N > K$. The analogous case where instead $\mathbf{Q}$ is known is omitted.

Now, it is clear that any positive definite weighting matrices can be used to provide consistent, but not efficient, preliminary estimates of $\boldsymbol{\omega}, \boldsymbol{\theta}$ and $\mathbf{b}$ [Eri02]. These are then used to yield consistent estimates of the unknown noise properties, e.g. by (3.44). Finally, efficient signal parameter estimates are obtained by again minimizing (3.46), but this time using the approximately optimal weighting. According to [Eri02], a reasonable choice of spatial weighting, especially for low signal to noise ratios, is

$$\mathbf{W}_Q = \left(\hat{\mathbf{R}} + \alpha\mathbf{I}\right)^{-1} , \tag{3.47}$$

where $\hat{\mathbf{R}}$ is the sample spatial covariance matrix of the data, defined by

$$\hat{\mathbf{R}} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H , \tag{3.48}$$

and $\alpha\mathbf{I}$, α relatively small, is a regularization which makes $\hat{\mathbf{R}}$ better conditioned. The regularization technique is in the array processing literature

often referred to as "diagonal loading", [Car88]. If this matrix is used as weighting, the resulting estimates will "almost" be asymptotically efficient for low to moderate signal to noise ratios. For very high signal to noise ratios, when $\hat{\mathbf{R}}$ significantly deviates from $\mathbf{Q}$, care has to be taken so that the use of $\mathbf{W}_Q$ as defined in Equation (3.47) does not result in a cancellation of the interesting source signals [Cox73]. This, together with a variance reduction, is the motivation for employing regularization.

Let us now elaborate on the minimization of (3.46). The WLS problem is brought back to the unweighted LS criterion (3.33) by absorbing the weighting matrices into the data and the signal model. Defining

$$\begin{aligned} \mathbf{Y}_W &= \mathbf{W}_Q^{1/2}\mathbf{Y}\mathbf{W}_C^{1/2} && (3.49)\\ \mathbf{A}_W &= \mathbf{W}_Q^{1/2}\mathbf{A} && (3.50)\\ \mathbf{S}_W &= \mathbf{S}\mathbf{W}_C^{1/2} && (3.51)\\ & && (3.52) \end{aligned}$$

results in the formulation

$$V_{\mathrm{WLS}}(\boldsymbol{\omega}, \boldsymbol{\theta}, \mathbf{b}) = \|\mathbf{Y}_W - \mathbf{A}_W\mathbf{B}\mathbf{S}_W\|_F^2. \tag{3.53}$$

Therefore, the signal parameter estimates are obtained similar to (3.37) and (3.40) as

$$\begin{aligned} \hat{\mathbf{b}} &= \left(\boldsymbol{\Upsilon}_W^H\boldsymbol{\Upsilon}_W\right)^{-1}\boldsymbol{\Upsilon}_W^H\mathbf{y}_{NW} && (3.54)\\ \left\{\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\theta}}\right\} &= \arg\min_{\boldsymbol{\omega},\boldsymbol{\theta}} \left\|\mathbf{P}_{\hat{\Upsilon}_W}^{\perp}(\boldsymbol{\omega}, \boldsymbol{\theta})\mathbf{y}_{NW}\right\|_F^2 && (3.55)\\ \mathbf{y}_{NW} &= \mathrm{vec}\left(\mathbf{Y}_W\right) && (3.56)\\ \boldsymbol{\Upsilon}_W &= \mathbf{S}_W^T \diamond \mathbf{A}_W\,. && (3.57) \end{aligned}$$

The 2-D WLS approach is now summarized as follows:

1. Solve (3.55) using column-weighting (3.47) and row-weighting $\mathbf{W}_C = \mathbf{C}^{-T}$.

2. Using the signal parameter estimates from Step 1, compute (3.54) and then (3.44).

3. Solve (3.55) again, using $\mathbf{W}_Q = \hat{\mathbf{Q}}^{-1}$.

3.6 Performance

The methods considered so far exploit the full 2-D data model, leading to high computational complexity. The resulting estimation accuracy is therefore of great interest. Unfortunately, exact expressions for the estimation error variance are not known, and would be overwhelmingly complicated. The standard approach is then to resort to an asymptotic analysis of the estimators. We will only consider the case of large N, the dual case of large K being similar. We will use a result from [Eri02] where consistency and the asymptotic distribution of the 2-D WLS estimator were derived under the following assumptions:

1. Each source is described by a unique $\{\theta, \omega\}$ pair.
2. The number of sources satisfies $M \leq K/2$.
3. The sequence of noise samples are zero mean, temporally independent, i.e. $\mathbf{C} = \mathbf{I}$, and have bounded fourth order moments.
4. The temporal sampling is uniform, i.e. $t = 1, \ldots, N$.

In the case of Gaussian distributed noise and a weighting matrix given by $\mathbf{W}_Q = \hat{\mathbf{Q}}^{-1}$, where $\hat{\mathbf{Q}}$ is a consistent estimate of $\mathbf{Q}$, the estimator was shown to be asymptotically efficient as $N \to \infty$. We can then make use of Theorem 3.4 to obtain a simple, explicit expression for the variance of the asymptotic distribution of the estimation error. The result is

$$\mathrm{E}[(\hat{\theta}_m - \theta_m)^2] = \frac{1}{2N|b|^2 \mathbf{d}_\theta^H \mathbf{Q}^{-1/2} \mathbf{P}^\perp_{\mathbf{Q}^{-1/2}\mathbf{a}} \mathbf{Q}^{-1/2} \mathbf{d}_\theta} \tag{3.58}$$

$$\mathrm{E}[(\hat{\omega}_m - \omega_m)^2] = \frac{6}{N^3 \, |b|^2 \, \mathbf{d}_\omega^H \mathbf{P}^\perp_{\mathbf{s}_N} \mathbf{d}_\omega \; \mathbf{a}^H \mathbf{Q}^{-1} \mathbf{a}} \tag{3.59}$$

(evaluated at b_m, ω_m and θ_m).

3A Derivation of the 2-D CRB

In this section, we derive the CRB on the DOA and Doppler parameters using the full 2-D model in 2.1. To this end, we use the general CRB for parameterized signals given in Theorem 3.1. The parameter vector is given by

$$\boldsymbol{\eta} = [\ \bar{\mathbf{b}}^T \quad \tilde{\mathbf{b}}^T \quad \boldsymbol{\omega}^T \quad \boldsymbol{\theta}^T\ ]^T \tag{3A.1}$$

where $\bar{\mathbf{b}} \triangleq \mathrm{Re}[\mathbf{b}]$ and $\tilde{\mathbf{b}} \triangleq \mathrm{Im}[\mathbf{b}]$. From Theorem 3.1 the FIM can be written as

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{\bar{b}\bar{b}} & \mathbf{J}_{\bar{b}\tilde{b}} & \mathbf{J}_{\bar{b}\omega} & \mathbf{J}_{\bar{b}\theta} \\ \mathbf{J}_{\bar{b}\tilde{b}}^T & \mathbf{J}_{\tilde{b}\tilde{b}} & \mathbf{J}_{\tilde{b}\omega} & \mathbf{J}_{\tilde{b}\theta} \\ \mathbf{J}_{\bar{b}\omega}^T & \mathbf{J}_{\tilde{b}\omega}^T & \mathbf{J}_{\omega\omega} & \mathbf{J}_{\omega\theta} \\ \mathbf{J}_{\bar{b}\theta}^T & \mathbf{J}_{\tilde{b}\theta}^T & \mathbf{J}_{\omega\theta}^T & \mathbf{J}_{\theta\theta} \end{bmatrix}, \tag{3A.2}$$

where

$$\begin{aligned}
\mathbf{J}_{\bar{b}\bar{b}} &= 2\mathrm{Re}\left[(\mathbf{S}_N^H\mathbf{C}^{-1}\mathbf{S}_N) \odot (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})\right] && \text{(3A.3)}\\
\mathbf{J}_{\bar{b}\tilde{b}} &= -2\mathrm{Im}\left[(\mathbf{S}_N^H\mathbf{C}^{-1}\mathbf{S}_N) \odot (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})\right] && \text{(3A.4)}\\
\mathbf{J}_{\bar{b}\omega} &= 2\mathrm{Re}\left[(\mathbf{S}_N^H\mathbf{C}^{-1}\mathbf{D}_\omega\mathbf{B}) \odot (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})\right] && \text{(3A.5)}\\
\mathbf{J}_{\bar{b}\theta} &= 2\mathrm{Re}\left[(\mathbf{S}_N^H\mathbf{C}^{-1}\mathbf{S}_N\mathbf{B}) \odot (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{D}_\theta)\right] && \text{(3A.6)}\\
\mathbf{J}_{\tilde{b}\tilde{b}} &= 2\mathrm{Re}\left[(\mathbf{S}_N^H\mathbf{C}^{-1}\mathbf{S}_N) \odot (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})\right] && \text{(3A.7)}\\
\mathbf{J}_{\tilde{b}\omega} &= 2\mathrm{Im}\left[(\mathbf{S}_N^H\mathbf{C}^{-1}\mathbf{D}_\omega\mathbf{B}) \odot (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})\right] && \text{(3A.8)}\\
\mathbf{J}_{\tilde{b}\theta} &= 2\mathrm{Im}\left[(\mathbf{S}_N^H\mathbf{C}^{-1}\mathbf{S}_N\mathbf{B}) \odot (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{D}_\theta)\right] && \text{(3A.9)}\\
\mathbf{J}_{\omega\omega} &= 2\mathrm{Re}\left[(\mathbf{B}^H\mathbf{D}_\omega^H\mathbf{C}^{-1}\mathbf{D}_\omega\mathbf{B}) \odot (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{A})\right] && \text{(3A.10)}\\
\mathbf{J}_{\omega\theta} &= 2\mathrm{Re}\left[(\mathbf{B}^H\mathbf{D}_\omega^H\mathbf{C}^{-1}\mathbf{S}_N\mathbf{B}) \odot (\mathbf{A}^H\mathbf{Q}^{-1}\mathbf{D}_\theta)\right] && \text{(3A.11)}\\
\mathbf{J}_{\theta\theta} &= 2\mathrm{Re}\left[(\mathbf{B}^H\mathbf{S}_N^H\mathbf{C}^{-1}\mathbf{S}_N\mathbf{B}) \odot (\mathbf{D}_\theta^H\mathbf{Q}^{-1}\mathbf{D}_\theta)\right]. && \text{(3A.12)}
\end{aligned}$$

Define

$$\mathbf{J}_{bb} \triangleq \begin{bmatrix} \mathbf{J}_{\bar{b}\bar{b}} & \mathbf{J}_{\bar{b}\tilde{b}} \\ \mathbf{J}_{\bar{b}\tilde{b}}^T & \mathbf{J}_{\tilde{b}\tilde{b}} \end{bmatrix}, \quad \mathbf{J}_1 \triangleq \begin{bmatrix} \mathbf{J}_{\bar{b}\omega} & \mathbf{J}_{\bar{b}\theta} \\ \mathbf{J}_{\tilde{b}\omega} & \mathbf{J}_{\tilde{b}\theta} \end{bmatrix}, \quad \mathbf{J}_2 \triangleq \begin{bmatrix} \mathbf{J}_{\omega\omega} & \mathbf{J}_{\omega\theta} \\ \mathbf{J}_{\omega\theta}^T & \mathbf{J}_{\theta\theta} \end{bmatrix}. \tag{3A.13}$$

The FIM can then be written as

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{bb} & \mathbf{J}_1 \\ \mathbf{J}_1^T & \mathbf{J}_2 \end{bmatrix}. \tag{3A.14}$$

We are interested in the CRB matrix for the DOAs and Doppler frequencies. This is given by the lower right $2M \times 2M$ block of the inverse of the FIM. Using a rule for the inverse of a partitioned matrix this is given by

$$\begin{bmatrix} \mathbf{C}_{\omega\omega} & \mathbf{C}_{\omega\theta} \\ \mathbf{C}_{\omega\theta}^T & \mathbf{C}_{\theta\theta} \end{bmatrix} = \left(\mathbf{J}_2 - \mathbf{J}_1^T \mathbf{J}_{bb}^{-1} \mathbf{J}_1\right)^{-1}. \tag{3A.15}$$

Now, define the following

$$\begin{aligned}
\mathbf{E} &\triangleq 2(\mathbf{S}_N^H \mathbf{C}^{-1} \mathbf{D}_\omega \mathbf{B}) \odot (\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A}) = \bar{\mathbf{E}} + j\tilde{\mathbf{E}} && \text{(3A.16)} \\
\mathbf{F} &\triangleq 2(\mathbf{S}_N^H \mathbf{C}^{-1} \mathbf{S}_N \mathbf{B}) \odot (\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{D}_\theta) = \bar{\mathbf{F}} + j\tilde{\mathbf{F}} && \text{(3A.17)} \\
\mathbf{H} &\triangleq 2(\mathbf{S}_N^H \mathbf{C}^{-1} \mathbf{S}_N) \odot (\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A}) = \bar{\mathbf{H}} + j\tilde{\mathbf{H}} && \text{(3A.18)} \\
\mathbf{G} &\triangleq \mathbf{H}^{-1} = \bar{\mathbf{G}} + j\tilde{\mathbf{G}}. && \text{(3A.19)}
\end{aligned}$$

We then have

$$\mathbf{J}_{bb} = \begin{bmatrix} \bar{\mathbf{H}} & -\tilde{\mathbf{H}} \\ \tilde{\mathbf{H}} & \bar{\mathbf{H}} \end{bmatrix}, \quad \mathbf{J}_1 = \begin{bmatrix} \bar{\mathbf{E}} & \bar{\mathbf{F}} \\ \tilde{\mathbf{E}} & \tilde{\mathbf{F}} \end{bmatrix}. \tag{3A.20}$$

From Appendix E, result R4 in [SN89] it follows that

$$\mathbf{J}_{bb}^{-1} = \begin{bmatrix} \bar{\mathbf{G}} & -\tilde{\mathbf{G}} \\ \tilde{\mathbf{G}} & \bar{\mathbf{G}} \end{bmatrix}. \tag{3A.21}$$

Some straightforward calculations then lead to

$$\mathbf{J}_1^T \mathbf{J}_{bb}^{-1} \mathbf{J}_1 = \mathrm{Re}\left\{ \begin{bmatrix} \mathbf{E}^H \mathbf{H}^{-1} \mathbf{E} & \mathbf{E}^H \mathbf{H}^{-1} \mathbf{F} \\ (\mathbf{E}^H \mathbf{H}^{-1} \mathbf{F})^T & \mathbf{F}^H \mathbf{H}^{-1} \mathbf{F} \end{bmatrix} \right\}. \tag{3A.22}$$

Substituting into (3A.15) we obtain

$$\begin{bmatrix} \mathbf{C}_{\omega\omega} & \mathbf{C}_{\omega\theta} \\ \mathbf{C}_{\omega\theta}^T & \mathbf{C}_{\theta\theta} \end{bmatrix} = \begin{bmatrix} \mathbf{J}_{\omega\omega} - \mathrm{Re}[\mathbf{E}^H \mathbf{H}^{-1} \mathbf{E}] & \mathbf{J}_{\omega\theta} - \mathrm{Re}[\mathbf{E}^H \mathbf{H}^{-1} \mathbf{F}] \\ \mathbf{J}_{\omega\theta}^T - \mathrm{Re}[\mathbf{E}^H \mathbf{H}^{-1} \mathbf{F}]^T & \mathbf{J}_{\theta\theta} - \mathrm{Re}[\mathbf{F}^H \mathbf{H}^{-1} \mathbf{F}] \end{bmatrix}^{-1}. \tag{3A.23}$$

Defining

$$\begin{aligned}
\mathbf{\Phi} &\triangleq \mathbf{J}_{\omega\omega} - \mathrm{Re}[\mathbf{E}^H \mathbf{H}^{-1} \mathbf{E}] && \text{(3A.24)} \\
\mathbf{\Psi} &\triangleq \mathbf{J}_{\theta\theta} - \mathrm{Re}[\mathbf{F}^H \mathbf{H}^{-1} \mathbf{F}] && \text{(3A.25)} \\
\mathbf{\Delta} &\triangleq \mathbf{J}_{\omega\theta} - \mathrm{Re}[\mathbf{E}^H \mathbf{H}^{-1} \mathbf{F}], && \text{(3A.26)}
\end{aligned}$$

and using standard rules for the inverse of a partitioned matrix, we obtain

$$\begin{aligned}
\mathbf{C}_{\omega\omega} &= \left(\mathbf{\Phi} - \mathbf{\Delta}\mathbf{\Psi}^{-1}\mathbf{\Delta}^T\right)^{-1} && \text{(3A.27)}\\
\mathbf{C}_{\theta\theta} &= \left(\mathbf{\Psi} - \mathbf{\Delta}^T\mathbf{\Phi}^{-1}\mathbf{\Delta}\right)^{-1} && \text{(3A.28)}\\
\mathbf{C}_{\omega\theta} &= -\left(\mathbf{\Phi} - \mathbf{\Delta}\mathbf{\Psi}^{-1}\mathbf{\Delta}^T\right)^{-1}\mathbf{\Delta}\mathbf{\Psi}^{-1}. && \text{(3A.29)}
\end{aligned}$$

To simplify these expressions, define

$$\begin{aligned}
\mathbf{\Upsilon} &= (\mathbf{C}^{-1/2}\mathbf{S}_N) \diamond (\mathbf{Q}^{-1/2}\mathbf{A}) && \text{(3A.30)}\\
\mathbf{\Upsilon}_\omega &= (\mathbf{C}^{-1/2}\mathbf{D}_\omega) \diamond (\mathbf{Q}^{-1/2}\mathbf{A}) && \text{(3A.31)}\\
\mathbf{\Upsilon}_\theta &= (\mathbf{C}^{-1/2}\mathbf{S}_N) \diamond (\mathbf{Q}^{-1/2}\mathbf{D}_\theta), && \text{(3A.32)}
\end{aligned}$$

and note that

$$(\mathbf{U}^H\mathbf{W}) \odot (\mathbf{V}^H\mathbf{Z}) = (\mathbf{U} \diamond \mathbf{V})^H (\mathbf{W} \diamond \mathbf{Z}) \qquad \text{(3A.33)}$$

for matrices with compatible dimensions. Substituting the definitions in (3A.30)–(3A.32) into (3A.16)–(3A.18) and using (3A.33) we obtain

$$\mathbf{E} = 2\mathbf{\Upsilon}^H\mathbf{\Upsilon}_\omega\mathbf{B}, \quad \mathbf{F} = 2\mathbf{\Upsilon}^H\mathbf{\Upsilon}_\theta\mathbf{B}, \quad \mathbf{H} = 2\mathbf{\Upsilon}^H\mathbf{\Upsilon}. \qquad \text{(3A.34)}$$

Similarly,

$$\begin{aligned}
\mathbf{J}_{\omega\omega} &= 2\mathrm{Re}\left[\mathbf{B}^H\mathbf{\Upsilon}_\omega^H\mathbf{\Upsilon}_\omega\mathbf{B}\right] && \text{(3A.35)}\\
\mathbf{J}_{\theta\theta} &= 2\mathrm{Re}\left[\mathbf{B}^H\mathbf{\Upsilon}_\theta^H\mathbf{\Upsilon}_\theta\mathbf{B}\right] && \text{(3A.36)}\\
\mathbf{J}_{\omega\theta} &= 2\mathrm{Re}\left[\mathbf{B}^H\mathbf{\Upsilon}_\omega^H\mathbf{\Upsilon}_\theta\mathbf{B}\right]. && \text{(3A.37)}
\end{aligned}$$

Inserting (3A.34)–(3A.37) into (3A.24)–(3A.26) then gives

$$\begin{aligned}
\mathbf{\Phi} &= 2\mathrm{Re}\left[\mathbf{B}^H\mathbf{\Upsilon}_\omega^H\mathbf{P}_\Upsilon^\perp\mathbf{\Upsilon}_\omega\mathbf{B}\right] && \text{(3A.38)}\\
\mathbf{\Psi} &= 2\mathrm{Re}\left[\mathbf{B}^H\mathbf{\Upsilon}_\theta^H\mathbf{P}_\Upsilon^\perp\mathbf{\Upsilon}_\theta\mathbf{B}\right] && \text{(3A.39)}\\
\mathbf{\Delta} &= 2\mathrm{Re}\left[\mathbf{B}^H\mathbf{\Upsilon}_\omega^H\mathbf{P}_\Upsilon^\perp\mathbf{\Upsilon}_\theta\mathbf{B}\right] && \text{(3A.40)}
\end{aligned}$$

where

$$\mathbf{P}_\Upsilon^\perp = \mathbf{I} - \mathbf{\Upsilon}(\mathbf{\Upsilon}^H\mathbf{\Upsilon})^{-1}\mathbf{\Upsilon}^H. \qquad \text{(3A.41)}$$

The CRB matrices are then obtained by substituting (3A.38)–(3A.40) into (3A.27)–(3A.29). Compact expressions can be obtained if we define

$$\mathbf{R}_\omega = \begin{bmatrix} \mathrm{Re}[\mathbf{P}_\Upsilon^\perp\mathbf{\Upsilon}_\omega\mathbf{B}] \\ \mathrm{Im}[\mathbf{P}_\Upsilon^\perp\mathbf{\Upsilon}_\omega\mathbf{B}] \end{bmatrix}, \quad \mathbf{R}_\theta = \begin{bmatrix} \mathrm{Re}[\mathbf{P}_\Upsilon^\perp\mathbf{\Upsilon}_\theta\mathbf{B}] \\ \mathrm{Im}[\mathbf{P}_\Upsilon^\perp\mathbf{\Upsilon}_\theta\mathbf{B}] \end{bmatrix}. \qquad \text{(3A.42)}$$

This leads to

$$\begin{aligned}\mathbf{C}_{\omega\omega} &= \left[2\left(\mathbf{R}_\omega^T\mathbf{R}_\omega - \mathbf{R}_\omega^T\mathbf{R}_\theta(\mathbf{R}_\theta^T\mathbf{R}_\theta)^{-1}\mathbf{R}_\theta^T\mathbf{R}_\omega\right)\right]^{-1} \\ &= \left(2\mathbf{R}_\omega^T\mathbf{P}_{\mathbf{R}_\theta}^{\perp}\mathbf{R}_\omega\right)^{-1} \qquad \text{(3A.43)}\end{aligned}$$

where $\mathbf{P}_{\mathbf{R}_\theta}^{\perp} = \mathbf{I} - \mathbf{R}_\theta(\mathbf{R}_\theta^T\mathbf{R}_\theta)^{-1}\mathbf{R}_\theta^T$. Similarly,

$$\mathbf{C}_{\theta\theta} = \left(2\mathbf{R}_\theta^T\mathbf{P}_{\mathbf{R}_\omega}^{\perp}\mathbf{R}_\theta\right)^{-1} \qquad \text{(3A.44)}$$

$$\mathbf{C}_{\omega\theta} = -\left(2\mathbf{R}_\omega^T\mathbf{P}_{\mathbf{R}_\theta}^{\perp}\mathbf{R}_\omega\right)^{-1}(\mathbf{R}_\theta^{\dagger}\mathbf{R}_\omega)^T. \qquad \text{(3A.45)}$$

This completes the proof of Theorem 3.2.

3B Derivation of the Asymptotic 2-D CRBs

In this section, we derive the asymptotic CRBs given in Theorem 3.3 and 3.4 respectively. In order to obtain a well defined CRB in the limit a proper normalization of the CRB matrix is crucial. We start with the case of large number of antennas and then give the case of large number of pulses.

Large Number of Antenna Elements

Starting from (3A.27)-(3A.28) we derive the asymptotic normalized CRBs

$$\begin{aligned}\bar{\mathbf{C}}_{\theta\theta} &= \lim_{K\to\infty} K^3\mathbf{C}_{\theta\theta} = \lim_{K\to\infty}\left[\frac{1}{K^3}\mathbf{\Psi} - \frac{1}{K^2}\mathbf{\Delta}^T\left(\frac{1}{K}\mathbf{\Phi}\right)^{-1}\frac{1}{K^2}\mathbf{\Delta}\right]^{-1} \\ &= \left(\bar{\mathbf{\Psi}} - \bar{\mathbf{\Delta}}^T\bar{\mathbf{\Phi}}^{-1}\bar{\mathbf{\Delta}}\right)^{-1}, \qquad \text{(3B.1)}\end{aligned}$$

and

$$\begin{aligned}\bar{\mathbf{C}}_{\omega\omega} &= \lim_{K\to\infty} K\mathbf{C}_{\omega\omega} = \lim_{K\to\infty}\left[\frac{1}{K}\mathbf{\Phi} - \frac{1}{K^2}\mathbf{\Delta}\left(\frac{1}{K^3}\mathbf{\Psi}\right)^{-1}\frac{1}{K^2}\mathbf{\Delta}^T\right]^{-1} \\ &= \left(\bar{\mathbf{\Phi}} - \bar{\mathbf{\Delta}}\bar{\mathbf{\Psi}}^{-1}\bar{\mathbf{\Delta}}^T\right)^{-1}, \qquad \text{(3B.2)}\end{aligned}$$

where

$$\bar{\mathbf{\Phi}} = \lim_{K\to\infty}\frac{1}{K}\mathbf{\Phi}, \qquad \bar{\mathbf{\Psi}} = \lim_{K\to\infty}\frac{1}{K^3}\mathbf{\Psi}, \qquad \bar{\mathbf{\Delta}} = \lim_{K\to\infty}\frac{1}{K^2}\mathbf{\Delta}. \qquad \text{(3B.3)}$$

According to the assumptions, all the asymptotic normalized matrices given above exist and are diagonal. For example, from the definition above and (3A.24), $\bar{\mathbf{\Phi}}$ is given by

$$\begin{aligned}\bar{\mathbf{\Phi}} &= \lim_{K\to\infty} 2\mathrm{Re}\Bigg[(\mathbf{B}^H\mathbf{D}_\omega^H\mathbf{C}^{-1}\mathbf{D}_\omega\mathbf{B})\odot(\frac{1}{K}\mathbf{A}^H\mathbf{A}) \\ &- \left((\mathbf{B}^H\mathbf{D}_\omega^H\mathbf{C}^{-1}\mathbf{S}_N)\odot(\frac{1}{K}\mathbf{A}^H\mathbf{A})\right)\left((\mathbf{S}_N^H\mathbf{C}^{-1}\mathbf{S}_N)\odot(\frac{1}{K}\mathbf{A}^H\mathbf{A})\right)^{-1} \\ &\times \left((\mathbf{S}_N^H\mathbf{C}^{-1}\mathbf{D}_\omega\mathbf{B})\odot(\frac{1}{K}\mathbf{A}^H\mathbf{A})\right)\Bigg]. \end{aligned} \tag{3B.4}$$

According to the assumptions, $\lim_{K\to\infty}\frac{1}{K}\mathbf{A}^H\mathbf{A}$ is diagonal and, due to the Hadamard products in (3B.4), $\bar{\mathbf{\Phi}}$ will be diagonal. The matrix operations in (3B.4) can thus be replaced by their scalar counterparts. The m-th diagonal element in $\bar{\mathbf{\Phi}}$ is then easily found as

$$\begin{aligned}\{\bar{\mathbf{\Phi}}\}_{mm} &= \lim_{K\to\infty}\frac{1}{K}2\mathrm{Re}\left[|b|^2\mathbf{d}_\omega^H\mathbf{C}^{-1}\mathbf{d}_\omega\mathbf{a}^H\mathbf{a} - \frac{|b|^2\mathbf{d}_\omega^H\mathbf{C}^{-1}\mathbf{s}_N\mathbf{a}^H\mathbf{a}\mathbf{s}_N^H\mathbf{C}^{-1}\mathbf{d}_\omega\mathbf{a}^H\mathbf{a}}{\mathbf{s}_N^H\mathbf{C}^{-1}\mathbf{s}_N\mathbf{a}^H\mathbf{a}}\right] \\ &= \lim_{K\to\infty}\frac{1}{K}2|b|^2\mathbf{a}^H\mathbf{a}\mathbf{d}_\omega^H\mathbf{C}^{-1/2}\left(\mathbf{I}-\frac{\mathbf{C}^{-1/2}\mathbf{s}_N\mathbf{s}_N^H\mathbf{C}^{-1/2}}{\mathbf{s}_N^H\mathbf{C}^{-1}\mathbf{s}_N}\right)\mathbf{C}^{-1/2}\mathbf{d}_\omega \\ &= \lim_{K\to\infty}\frac{1}{K}2|b|^2\mathbf{a}^H\mathbf{a}\mathbf{d}_\omega^H\mathbf{C}^{-1/2}\mathbf{P}^\perp_{\mathbf{C}^{-1/2}\mathbf{s}_N}\mathbf{C}^{-1/2}\mathbf{d}_\omega\,, \end{aligned} \tag{3B.5}$$

where the arguments of $b_m, \mathbf{d}_\omega(\omega_m), \mathbf{a}(\theta_m)$ and $\mathbf{s}_N(\omega_m)$ have been suppressed for brevity. By similar arguments, $\bar{\mathbf{\Psi}}$ and $\bar{\mathbf{\Delta}}$ are also diagonal, with diagonal elements given by

$$\{\bar{\mathbf{\Psi}}\}_{mm} = \lim_{K\to\infty}\frac{1}{K^3}2|b|^2\mathbf{s}_N^H\mathbf{C}^{-1}\mathbf{s}_N\mathbf{d}_\theta^H\mathbf{P}_\mathbf{a}^\perp\mathbf{d}_\theta \tag{3B.6}$$

$$\begin{aligned}\{\bar{\mathbf{\Delta}}\}_{mm} &= \lim_{K\to\infty}\frac{1}{K^2}2\mathrm{Re}\left[|b|^2\mathbf{d}_\omega^H\mathbf{C}^{-1}\mathbf{s}_N\mathbf{a}^H\mathbf{d}_\theta - |b|^2\mathbf{d}_\omega^H\mathbf{C}^{-1}\mathbf{s}_N\mathbf{a}^H\mathbf{d}_\theta\right] \\ &= 0. \end{aligned} \tag{3B.7}$$

Thus, we obtain

$$\bar{\mathbf{C}}_{\theta\theta} = \bar{\mathbf{\Psi}}^{-1}, \quad \bar{\mathbf{C}}_{\omega\omega} = \bar{\mathbf{\Phi}}^{-1}, \quad \text{and} \quad \lim_{K\to\infty}\mathbf{C}_{\omega\theta} = \mathbf{0}. \tag{3B.8}$$

The m-th diagonal elements in $\bar{\mathbf{C}}_{\theta\theta}$ and $\bar{\mathbf{C}}_{\omega\omega}$ are given by

$$\{\bar{\mathbf{C}}_{\theta\theta}\}_{mm} = \lim_{K\to\infty} K^3\frac{1}{2|b|^2\mathbf{s}_N^H\mathbf{C}^{-1}\mathbf{s}_N\mathbf{d}_\theta^H\mathbf{P}_\mathbf{a}^\perp\mathbf{d}_\theta} \tag{3B.9}$$

$$\{\bar{\mathbf{C}}_{\omega\omega}\}_{mm} = \lim_{K\to\infty} K\frac{1}{2|b|^2\mathbf{a}^H\mathbf{a}\mathbf{d}_\omega^H\mathbf{C}^{-1/2}\mathbf{P}^\perp_{\mathbf{C}^{-1/2}\mathbf{s}_N}\mathbf{C}^{-1/2}\mathbf{d}_\omega}. \tag{3B.10}$$

This means that for sufficiently large K, the DOA and Doppler CRBs are given by

$$c_\theta = \frac{1}{2|b|^2 \mathbf{s}_N^H \mathbf{C}^{-1} \mathbf{s}_N \mathbf{d}_\theta^H \mathbf{P}_{\mathbf{a}}^{\perp} \mathbf{d}_\theta} \tag{3B.11}$$

$$c_\omega = \frac{1}{2|b|^2 \mathbf{a}^H \mathbf{a} \mathbf{d}_\omega^H \mathbf{C}^{-1/2} \mathbf{P}_{\mathbf{C}^{-1/2}\mathbf{s}_N}^{\perp} \mathbf{C}^{-1/2} \mathbf{d}_\omega}. \tag{3B.12}$$

Clearly, the single-source bound will be the same as this asymptotic bound since a diagonal element in the diagonal matrices above is the same as if only one source was present.

Large Number of Pulses

When $N \to \infty$, different normalizations of the asymptotic CRB matrices are required. Thus we derive the following asymptotic normalized CRBs

$$\begin{aligned} \bar{\mathbf{C}}_{\theta\theta} &= \lim_{N\to\infty} N\mathbf{C}_{\theta\theta} = \lim_{N\to\infty} \left[\frac{1}{N}\mathbf{\Psi} - \frac{1}{N^2}\mathbf{\Delta}^T \left(\frac{1}{N^3}\mathbf{\Phi} \right)^{-1} \frac{1}{N^2}\mathbf{\Delta} \right]^{-1} \\ &= \left(\bar{\mathbf{\Psi}} - \bar{\mathbf{\Delta}}^T \bar{\mathbf{\Phi}}^{-1} \bar{\mathbf{\Delta}} \right)^{-1} \end{aligned} \tag{3B.13}$$

and

$$\begin{aligned} \bar{\mathbf{C}}_{\omega\omega} &= \lim_{N\to\infty} N^3\mathbf{C}_{\omega\omega} = \lim_{N\to\infty} \left[\frac{1}{N^3}\mathbf{\Phi} - \frac{1}{N^2}\mathbf{\Delta} \left(\frac{1}{N}\mathbf{\Psi} \right)^{-1} \frac{1}{N^2}\mathbf{\Delta}^T \right]^{-1} \\ &= \left(\bar{\mathbf{\Phi}} - \bar{\mathbf{\Delta}} \bar{\mathbf{\Psi}}^{-1} \bar{\mathbf{\Delta}}^T \right)^{-1} \end{aligned} \tag{3B.14}$$

where we now have the normalizations

$$\bar{\mathbf{\Phi}} = \lim_{N\to\infty} \frac{1}{N^3}\mathbf{\Phi}, \qquad \bar{\mathbf{\Psi}} = \lim_{N\to\infty} \frac{1}{N}\mathbf{\Psi}, \qquad \bar{\mathbf{\Delta}} = \lim_{N\to\infty} \frac{1}{N^2}\mathbf{\Delta}. \tag{3B.15}$$

By similar arguments as in the case of $K \to \infty$, the matrices above exist and are diagonal. Proceeding as in (3B.4)–(3B.7) yields (in this case, the first factors in the Hadamard products are diagonal)

$$\begin{aligned} \bar{\mathbf{\Phi}} &= \lim_{N\to\infty} 2\mathrm{Re}\Bigg[(\mathbf{B}^H \frac{1}{N^3} \mathbf{D}_\omega^H \mathbf{D}_\omega \mathbf{B}) \odot (\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A}) \\ &- \left((\mathbf{B}^H \frac{1}{N^2} \mathbf{D}_\omega^H \mathbf{S}_N) \odot (\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A}) \right) \left((\frac{1}{N} \mathbf{S}_N^H \mathbf{S}_N) \odot (\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A}) \right)^{-1} \\ &\times \left((\frac{1}{N^2} \mathbf{S}_N^H \mathbf{D}_\omega \mathbf{B}) \odot (\mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A}) \right) \Bigg], \end{aligned} \tag{3B.16}$$

with diagonal elements

$$\begin{aligned}\{\bar{\mathbf{\Phi}}\}_{mm} &= \lim_{N\to\infty}\frac{1}{N^3}2\mathrm{Re}\Big[|b|^2\mathbf{d}_\omega^H\mathbf{d}_\omega\mathbf{a}^H\mathbf{Q}^{-1}\mathbf{a} \\ &- \frac{|b|^2\mathbf{d}_\omega^H\mathbf{s}_N\mathbf{a}^H\mathbf{Q}^{-1}\mathbf{a}\mathbf{s}_N^H\mathbf{d}_\omega\mathbf{a}^H\mathbf{Q}^{-1}\mathbf{a}}{\mathbf{s}_N^H\mathbf{s}_N\mathbf{a}^H\mathbf{Q}^{-1}\mathbf{a}}\Big] \\ &= \lim_{N\to\infty}\frac{1}{N^3}2|b|^2\mathbf{d}_\omega^H\mathbf{P}^\perp_{\mathbf{s}_N}\mathbf{d}_\omega\ \mathbf{a}^H\mathbf{Q}^{-1}\mathbf{a}. \qquad (3B.17)\end{aligned}$$

Furthermore,

$$\begin{aligned}\{\bar{\mathbf{\Psi}}\}_{mm} &= \lim_{N\to\infty}\frac{1}{N}2\mathrm{Re}\Big[|b|^2\mathbf{s}_N^H\mathbf{s}_N\mathbf{d}_\theta^H\mathbf{Q}^{-1}\mathbf{d}_\theta \\ &- \frac{|b|^2\mathbf{s}_N^H\mathbf{s}_N\mathbf{d}_\theta^H\mathbf{Q}^{-1}\mathbf{a}\mathbf{s}_N^H\mathbf{s}_N\mathbf{a}^H\mathbf{Q}^{-1}\mathbf{d}_\theta}{\mathbf{s}_N^H\mathbf{s}_N\mathbf{a}^H\mathbf{Q}^{-1}\mathbf{a}}\Big] \\ &= \lim_{N\to\infty}\frac{1}{N}2|b|^2\ \mathbf{s}_N^H\mathbf{s}_N\ \mathbf{d}_\theta^H\mathbf{Q}^{-1/2}\mathbf{P}^\perp_{\mathbf{Q}^{-1/2}\mathbf{a}}\mathbf{Q}^{-1/2}\mathbf{d}_\theta, \qquad (3B.18) \\ \{\bar{\mathbf{\Delta}}\}_{mm} &= \lim_{N\to\infty}\frac{1}{N^2}2\mathrm{Re}\left[|b|^2\mathbf{d}_\omega^H\mathbf{s}_N\mathbf{a}^H\mathbf{Q}^{-1}\mathbf{d}_\theta - |b|^2\mathbf{d}_\omega^H\mathbf{s}_N\mathbf{a}^H\mathbf{Q}^{-1}\mathbf{d}_\theta\right] = 0. \\ & \qquad (3B.19)\end{aligned}$$

Thus, we obtain

$$\bar{\mathbf{C}}_{\theta\theta} = \bar{\mathbf{\Psi}}^{-1}, \quad \bar{\mathbf{C}}_{\omega\omega} = \bar{\mathbf{\Phi}}^{-1}, \quad \text{and} \quad \lim_{N\to\infty}\mathbf{C}_{\omega\theta} = \mathbf{0}. \qquad (3B.20)$$

The m-th diagonal elements in $\bar{\mathbf{C}}_{\theta\theta}$ and $\bar{\mathbf{C}}_{\omega\omega}$ are given by

$$\{\bar{\mathbf{C}}_{\theta\theta}\}_{mm} = \lim_{N\to\infty} N\frac{1}{2|b|^2\ \mathbf{s}_N^H\mathbf{s}_N\ \mathbf{d}_\theta^H\mathbf{Q}^{-1/2}\mathbf{P}^\perp_{\mathbf{Q}^{-1/2}\mathbf{a}}\mathbf{Q}^{-1/2}\mathbf{d}_\theta} \qquad (3B.21)$$

$$\{\bar{\mathbf{C}}_{\omega\omega}\}_{mm} = \lim_{N\to\infty} N^3\frac{1}{2|b|^2\mathbf{d}_\omega^H\mathbf{P}^\perp_{\mathbf{s}_N}\mathbf{d}_\omega\ \mathbf{a}^H\mathbf{Q}^{-1}\mathbf{a}}. \qquad (3B.22)$$

This means that for sufficiently large N the DOA and Doppler CRBs are given by

$$c_\theta = \frac{1}{2|b|^2\ \mathbf{s}_N^H\mathbf{s}_N\ \mathbf{d}_\theta^H\mathbf{Q}^{-1/2}\mathbf{P}^\perp_{\mathbf{Q}^{-1/2}\mathbf{a}}\mathbf{Q}^{-1/2}\mathbf{d}_\theta} \qquad (3B.23)$$

$$c_\omega = \frac{1}{2|b|^2\mathbf{d}_\omega^H\mathbf{P}^\perp_{\mathbf{s}_N}\mathbf{d}_\omega\ \mathbf{a}^H\mathbf{Q}^{-1}\mathbf{a}}. \qquad (3B.24)$$

Using the same argument as in the $K\to\infty$ case, this asymptotic bound will be the same as the single target bound. This completes the proof of Theorems 3.3 and 3.4. □

Chapter 4

Estimation Using a Decoupled 1-D/1-D Model

The underlying idea in the 1-D/1-D approach is to perform the 2-D estimation in two steps. This is possible by first relaxing the parameterization in one of the dimensions. Although this increases the number of free parameters, we shall see that the loss in estimation accuracy can be made negligible in large samples. The idea was first presented in [SS98] for the case of a single source, and we extend it here to multiple sources.

4.1 CRB for the 1-D Models

To begin with, we investigate the achievable performance, as dictated by the CRB, using the decoupled model.

Spatial Data Model

Relaxing the temporal signal parameterization, the 1-D DOA model is expressed as in (1.12)

$$\mathbf{y}(n) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(n) + \mathbf{n}(n), \qquad n = 1, \ldots, N\,. \tag{4.1}$$

The signal samples $\mathbf{x}(n),\ n = 1, \ldots, N$, are considered deterministic parameters to be estimated along with $\boldsymbol{\theta}$. Since the Doppler parameterization is not utilized, the DOA parameters are uniquely determined by (4.1) only if they are distinct. Also, the number of signals, M, must be less than the number of sensors, K. As before, the FIM is block diagonal, so that the CRB on $\boldsymbol{\theta}$ and $\mathbf{x}(n)$ does not depend on the parameterization of the noise

covariance. However, using the decoupled model (4.1) it is no longer possible to estimate the spatial covariance $\mathbf{Q}$. Thus, only $\mathbf{C}$ can be unknown in this case, and therefore $K > N$ is assumed. The CRB for the case of white noise was derived in [SN89]. The extension to colored noise is immediate using the pre-whitening approach, and results in

Theorem 4.1 *Let $\hat{\boldsymbol{\theta}}$ be an unbiased estimator of $\boldsymbol{\theta}$ in (4.1). Assume that $\mathbf{Q}$ is known, $M < K$, $N < K$ (if $\mathbf{C} \propto \mathbf{I}$ the latter condition is not needed); and that $\theta_m \neq \theta_l$ for $m \neq l$. Then, the CRB for the DOA parameters is given by*

$$\begin{aligned} &\mathrm{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] \geq \mathbf{C}_\theta \\ &\mathbf{C}_\theta = \left(2\mathrm{Re}\left[(\mathbf{X}_N^H \mathbf{C}^{-1} \mathbf{X}_N) \odot (\mathbf{D}_\theta^H \mathbf{Q}^{-1/2} \mathbf{P}^\perp_{\mathbf{Q}^{-1/2}\mathbf{A}} \mathbf{Q}^{-1/2} \mathbf{D}_\theta)\right]\right)^{-1}, \end{aligned} \tag{4.2}$$

where

$$\mathbf{X}_N = \begin{bmatrix} x_1(1) & \cdots & x_M(1) \\ \vdots & & \vdots \\ x_1(N) & \cdots & x_M(N) \end{bmatrix}. \tag{4.3}$$

Proof: If the noise parameterization is functionally independent of the signal parameters the FIM will be block-diagonal with respect to the signal and noise parameters. Therefore, the CRB on the signal parameters is the same whether the noise covariance is known or not. Without loss of generality, we can therefore assume that the noise covariance matrix is known. The theorem then follows by applying the CRB for the white noise case in [SN89] to the pre-whitened data model

$$\mathbf{Y}_W = \mathbf{Q}^{-1/2}\mathbf{Y}\mathbf{C}^{-T/2} = \mathbf{A}_W(\boldsymbol{\theta})\mathbf{X}_W + \mathbf{N}_W\,.$$

Temporal Data Model

Similar to (4.1), the unstructured 1-D Doppler model is given by

$$\mathbf{y}(n) = \boldsymbol{\Gamma}\mathbf{s}(\boldsymbol{\omega}, n) + \mathbf{n}(n), \qquad n = 1, \ldots, N\,. \tag{4.4}$$

where, now, the elements of the complex $K \times M$ matrix $\boldsymbol{\Gamma}$ are regarded as free parameters. By symmetry, we obtain the dual result of Theorem 4.1 as:

Theorem 4.2 *Let $\mathbf{C}$ be known and assume that $M < N$, $K < N$ (if $\mathbf{Q} \propto \mathbf{I}$ the latter condition is not needed); and that $\omega_m \neq \omega_l$ for $m \neq l$. The CRB for the Doppler parameters in the model (4.4) is then*

$$\begin{aligned} &\mathrm{E}[(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})^T] \geq \mathbf{C}_\omega \\ &\mathbf{C}_\omega = \left(2\mathrm{Re}\left[(\mathbf{D}_\omega^H \mathbf{C}^{-1/2} \mathbf{P}_{\mathbf{C}^{-1/2}\mathbf{S}_N} \mathbf{C}^{-1/2} \mathbf{D}_\omega) \odot (\boldsymbol{\Gamma}^H \mathbf{Q}^{-1} \boldsymbol{\Gamma})\right]\right)^{-1}. \end{aligned} \tag{4.5}$$

4.2 Asymptotic CRB for the 1-D Models

We now prove that, under some circumstances, the 2-D CRB can be attained using the decoupled 1-D/1-D models.

1-D DOA CRB, $K \to \infty$

In the spatial 1-D model, the signal matrix $\mathbf{X}$ is composed of free parameters that need to be estimated along with the DOA . Since $\mathbf{X}$ is $M \times N$, this matrix can only be consistently estimated as $K \to \infty$. Thus, this model is suitable for the case of large number of antennas:

Theorem 4.3 *Let the assumptions of Theorem 3.3 hold, and assume that the DOAs are distinct. Then, for $K \to \infty$, the decoupled DOA CRB (4.2) is identical to the asymptotic ($K \to \infty$) 2-D CRB given by (3.21).*

Proof: Follows from the fact that the matrix $\mathbf{D}_\theta^H \mathbf{P}_{\mathbf{A}}^\perp \mathbf{D}_\theta$, appearing in (4.2) (with $\mathbf{Q} = \mathbf{I}$), is asymptotically diagonal with elements $\mathbf{d}_\theta^H \mathbf{P}_{\mathbf{a}}^\perp \mathbf{d}_\theta$.

1-D Doppler CRB, $N \to \infty$

By the same arguments as above, the decoupled Doppler model (4.4) is suitable for the case of large number of temporal samples.

Theorem 4.4 *Let the assumptions of Theorem 3.4 hold, and assume that the ω_ms are distinct. Then, for $N \to \infty$, the decoupled Doppler CRB (4.5) is identical to the asymptotic ($N \to \infty$) 2-D CRB given by (3.24).*

The interesting conclusion from the above results is that 2-D optimal performance using the decoupled signal model might be possible, but the estimates must be obtained in the right order! More precisely, for large K one must determine $\hat{\boldsymbol{\theta}}$ and an unstructured estimate of $\mathbf{X}$ first. We shall see later that 2-D optimal estimation of $\hat{\boldsymbol{\omega}}$ is also possible, by fitting the temporal model $\mathbf{X} = \mathbf{BS}(\boldsymbol{\omega})$ to the preliminary estimate $\hat{\mathbf{X}}$. Similarly, when N is large it is necessary to instead estimate $\boldsymbol{\omega}$ in the first step. Optimal DOA estimates can then be obtained by fitting the spatial signal model to the unstructured estimate of $\boldsymbol{\Gamma}$.

We have seen that the single-source asymptotic 1-D and 2-D CRBs and the multi-source asymptotic 1-D and 2-D CRBs are all identical. Thus, there are two types of decoupling. As the amount of data tends to infinity in one of the frequency dimensions, both the sources and the frequency dimensions decouple.

4.3 1-D/1-D Weighted Least Squares

Having concluded that it is possible to obtain asymptotically 2-D optimal estimates using the decoupled signal model, we now search for means to achieve this performance. The derivation is performed for the Doppler model (4.4), repeated below for convenience. The corresponding method based on (4.1) is then directly obtained by symmetry. Now, consider the decoupled temporal signal model

$$\mathbf{y}(n) = \mathbf{\Gamma}\mathbf{s}(\boldsymbol{\omega}, n) + \mathbf{n}(n)\,, \tag{4.6}$$

where $\mathbf{n}(n)$ is assumed to be temporally white. The parsimonious parameterization of $\mathbf{\Gamma}$ would be

$$\mathbf{\Gamma} = \mathbf{A}(\boldsymbol{\theta})\mathbf{B} = \sum_{m=1}^{M} \mathbf{a}(\theta_m) b_m\,. \tag{4.7}$$

In the 1-D/1-D approach, $\boldsymbol{\omega}$ and $\mathbf{\Gamma}$ are first estimated ignoring the structure of $\mathbf{\Gamma}$. Then, the DOA parameters are found by imposing (4.7) to the preliminary estimate $\hat{\mathbf{\Gamma}}$ in a second step. Since the mapping from the parameters of the unstructured $\mathbf{\Gamma}$ (e.g. the real and imaginary parts of the matrix elements) to $\boldsymbol{\theta}$ and $\mathbf{b}$ is many-to-one, the resulting estimates of $\boldsymbol{\theta}$ and $\mathbf{b}$ are suboptimal. However, by employing the extended invariance principle (EXIP) of [SS89], we retain the asymptotic (for large N) efficiency.

If $\mathbf{n}(n)$ is temporally and spatially white, the exact ML estimate of $\mathbf{\Gamma}$ and $\boldsymbol{\omega}$ can be found by a nonlinear LS fit. Though the exact ML method also is tractable in the spatially colored noise case, here we will consider the asymptotically equivalent WLS approach. Thus, the unstructured spatial data matrix and the Doppler frequencies are found by solving

$$\{\hat{\boldsymbol{\omega}}, \hat{\mathbf{\Gamma}}\} = \arg\min_{\boldsymbol{\omega},\mathbf{\Gamma}} \mathrm{Tr}\big\{[\mathbf{Y} - \mathbf{\Gamma}\mathbf{S}(\boldsymbol{\omega})]^H \mathbf{W}_Q [\mathbf{Y} - \mathbf{\Gamma}\mathbf{S}(\boldsymbol{\omega})]\big\}\,, \tag{4.8}$$

where $\mathbf{W}_Q$ is the spatial weighting matrix. If $\mathbf{Q} = \mathrm{E}[\mathbf{n}(n)\mathbf{n}^H(n)]$ is known, (4.8) yields the exact ML estimates when $\mathbf{W}_Q = \mathbf{Q}^{-1}$ is employed. For the case of unknown noise color, we propose a two-step approach below similar to that of Section 3.5.

Absorbing the weighting matrix into $\mathbf{Y}$ and $\mathbf{\Gamma}$, (4.8) is conveniently rewritten as

$$\{\hat{\boldsymbol{\omega}}, \hat{\mathbf{\Gamma}}_W\} = \arg\min_{\boldsymbol{\omega},\mathbf{\Gamma}} \|\mathbf{Y}_W - \mathbf{\Gamma}_W \mathbf{S}(\boldsymbol{\omega})\|_F^2 \tag{4.9}$$

$$\mathbf{Y}_W = \mathbf{W}_Q^{1/2}\mathbf{Y} \tag{4.10}$$

$$\mathbf{\Gamma}_W = \mathbf{W}_Q^{1/2}\mathbf{\Gamma}\,. \tag{4.11}$$

For fixed $\boldsymbol{\omega}$, the WLS estimate of $\boldsymbol{\Gamma}$ is now found as

$$\hat{\boldsymbol{\Gamma}} = \mathbf{W}_Q^{-1/2}\hat{\boldsymbol{\Gamma}}_W = \mathbf{Y}\mathbf{S}^H(\boldsymbol{\omega})\left[\mathbf{S}(\boldsymbol{\omega})\mathbf{S}^H(\boldsymbol{\omega})\right]^{-1}. \tag{4.12}$$

Substituting (4.12) into (4.9), the Doppler frequencies are determined by solving

$$\hat{\boldsymbol{\omega}} = \arg\min_{\boldsymbol{\omega}} \left\| \mathbf{Y}_W - \hat{\boldsymbol{\Gamma}}_W\mathbf{S} \right\|_F^2 = \arg\min_{\boldsymbol{\omega}} \left\| \mathbf{W}_Q^{1/2}\mathbf{Y}\mathbf{P}^{\perp}_{\mathbf{S}^H} \right\|_F^2, \tag{4.13}$$

where $\mathbf{P}^{\perp}_{\mathbf{S}^H} = \mathbf{I} - \mathbf{S}^H\left(\mathbf{S}\mathbf{S}^H\right)^{-1}\mathbf{S}$. Although (4.13) is again a complicated nonlinear optimization, the dimension has been reduced from $2M$ parameters in (3.55) to M in (4.13).

It remains to determine the $\boldsymbol{\theta}$ and $\mathbf{b}$ parameters. According to the EXIP principle, these are again determined by WLS, with the weighting given by the FIM for the problem at hand. Similar to [SS98], where only the case of $M = 1$ was considered, the unstructured model is re-written as

$$\mathrm{vec}\left(\mathbf{Y}_W\right) = (\mathbf{S}^T \otimes \mathbf{I})\boldsymbol{\gamma} + \mathrm{vec}(\mathbf{N}), \tag{4.14}$$

where $\boldsymbol{\gamma} = \mathrm{vec}(\boldsymbol{\Gamma})$. We will consider this as a plain linear regression in $\boldsymbol{\gamma}$, ignoring the fact that the parameters $\boldsymbol{\omega}$ of $\mathbf{S}$ are also unknown. This is justified by the high accuracy of $\hat{\boldsymbol{\omega}}$ (variance proportional to $1/N^3$) as compared to that of $\hat{\boldsymbol{\Gamma}}$ (variance proportional to $1/N$). Applying [Kay93, Section 15.7], we derive the FIM for estimation of $\boldsymbol{\gamma}$ as

$$\mathbf{F}(\boldsymbol{\gamma}) = \left(\mathbf{S}\mathbf{S}^H\right)^T \otimes \mathbf{Q}^{-1} \simeq N\,\mathbf{I} \otimes \mathbf{Q}^{-1}, \tag{4.15}$$

where the approximation is valid for large N. The EXIP principle now dictates

$$\{\hat{\boldsymbol{\theta}}, \hat{\mathbf{b}}\} = \arg\min_{\boldsymbol{\theta},\boldsymbol{b}} \mathrm{vec}^H\left(\hat{\boldsymbol{\Gamma}} - \mathbf{A}(\boldsymbol{\theta})\mathbf{B}\right)\left(\mathbf{I} \otimes \mathbf{Q}^{-1}\right)\mathrm{vec}\left(\hat{\boldsymbol{\Gamma}} - \mathbf{A}(\boldsymbol{\theta})\mathbf{B}\right). \tag{4.16}$$

By the properties of the Kronecker product, this can be re-written as

$$\{\hat{\boldsymbol{\theta}}, \hat{\mathbf{b}}\} = \arg\min_{\boldsymbol{\theta},\boldsymbol{b}} \mathrm{Tr}\Big\{[\hat{\boldsymbol{\Gamma}} - \mathbf{A}(\boldsymbol{\theta})\mathbf{B}]^H\mathbf{W}_Q[\hat{\boldsymbol{\Gamma}} - \mathbf{A}(\boldsymbol{\theta})\mathbf{B}]\Big\}, \tag{4.17}$$

with $\mathbf{W}_Q = \mathbf{Q}^{-1}$. Using that $\mathbf{B}$ is diagonal results in

$$\{\hat{\boldsymbol{\theta}}, \hat{\mathbf{b}}\} = \arg\min_{\boldsymbol{\theta},\boldsymbol{b}} \sum_{m=1}^{M} [\hat{\boldsymbol{\Gamma}}_m - \mathbf{a}(\theta_m)b_m]^H\mathbf{W}_Q[\hat{\boldsymbol{\Gamma}}_m - \mathbf{a}(\theta_m)b_m], \tag{4.18}$$

where $\hat{\mathbf{\Gamma}}_m$ is the m-th column of $\hat{\mathbf{\Gamma}}$. Thus, the estimation of the M DOA parameters separates into M 1-D problems. Concentrating (4.18) with respect to b_m yields

$$\hat{b}_m = \frac{\mathbf{a}^H(\theta_m)\mathbf{W}_Q\hat{\mathbf{\Gamma}}_m}{\mathbf{a}^H(\theta_m)\mathbf{W}_Q\mathbf{a}(\theta_m)} \tag{4.19}$$

and, after straightforward manipulations,

$$\hat{\theta}_m = \arg\max_{\theta} \frac{|\mathbf{a}^H(\theta)\mathbf{W}_Q\hat{\mathbf{\Gamma}}_m|^2}{\mathbf{a}^H(\theta)\mathbf{W}_Q\mathbf{a}(\theta)}, \ m = 1, \ldots, M. \tag{4.20}$$

Hence, the m-th DOA is estimated by weighted beamforming, applied to the m-th column of $\hat{\mathbf{\Gamma}}$. Note that the 2-D problem has decoupled not only into two 1-D problems, but also that the M-D DOA search has decoupled into M 1-D searches. This is a tremendous reduction in computational complexity.

In the case of spatially white noise, or, equivalently, known $\mathbf{Q}$, the estimator is now completely specified. In the case of unknown $\mathbf{Q}$, the following 2-step approach is proposed:

1. Solve (4.13) using column-weighting (3.47).
2. Given preliminary $\boldsymbol{\omega}$ estimates, compute $\hat{\mathbf{\Gamma}}$ from (4.12), and update the spatial weighting by
$$\mathbf{W}_Q = \left(\frac{1}{N}(\mathbf{Y} - \hat{\mathbf{\Gamma}}\mathbf{S}(\hat{\boldsymbol{\omega}}))(\mathbf{Y} - \hat{\mathbf{\Gamma}}\mathbf{S}(\hat{\boldsymbol{\omega}}))^H\right)^{-1}. \tag{4.21}$$
3. Solve (4.13) again, using the asymptotically optimal weighting.
4. Compute (4.12) and (4.21) again, and solve (4.20) for the DOA parameters.

Given the significantly reduced complexity as compared to the full 2-D approach, the following result is remarkable.

Theorem 4.5 *Let $\hat{\boldsymbol{\omega}}$ and $\hat{\boldsymbol{\theta}}$ be determined by the above outlined 1-D/1-D estimator. Assume that the ω_ms are distinct and that the temporal sampling is uniform, i.e.*

$$\mathbf{s}(\boldsymbol{\omega}, t) \ = \ \left[\ e^{j\omega_1 t} \ \ \ldots \ \ e^{j\omega_M t} \ \right]^T, \quad t = 1, \ldots, N.$$

Assume further that the noise is temporally white and spatially colored with covariance matrix $\mathbf{Q}$. Assume that the array manifold is unambiguous and has bounded derivatives up to third order. Then, $\hat{\boldsymbol{\omega}}$ and $\hat{\boldsymbol{\theta}}$ are both consistent and asymptotically 2-D efficient as $N \to \infty$.

Proof: Follows essentially from the EXIP principle [SS89], but it must be proved that the estimation of $\boldsymbol{\omega}$ does not affect the FIM for $\boldsymbol{\gamma}$. We will not use the EXIP principle here to prove the theorem. Instead, we will use standard Taylor expansion techniques to derive the asymptotic properties of the proposed estimator. Since the proof is lengthy, the details are deferred to Appendix 4A.

The 1-D Doppler estimator

Consistency

To show consistency of the Doppler estimator in (4.13) express the criterion function as

$$J(\boldsymbol{\omega}) \triangleq \left\| \mathbf{W}^{1/2}\mathbf{Y}\mathbf{P}^{\perp}_{\mathbf{S}^H} \right\|_F^2 = \mathrm{Tr}\{\mathbf{Y}_W \mathbf{P}^{\perp}_{\mathbf{S}^H} \mathbf{Y}_W^H\}, \tag{4.22}$$

where $\mathbf{Y}_W = \mathbf{W}^{1/2}\mathbf{Y}$ and the Q subscript on the weighting matrix has been dropped for brevity. Introduce a row-wise slicing of the weighted data matrix according to

$$\mathbf{Y}_W = \begin{bmatrix} \breve{\mathbf{y}}_{W1} \\ \vdots \\ \breve{\mathbf{y}}_{WK} \end{bmatrix} \tag{4.23}$$

where the $1 \times N$ vector $\breve{\mathbf{y}}_{Wk}$ is the k-th row in $\mathbf{Y}_W$. The criterion function can then be written as

$$J(\boldsymbol{\omega}) = \sum_{k=1}^{K} \breve{\mathbf{y}}_{Wk} \mathbf{P}^{\perp}_{\mathbf{S}^H} \breve{\mathbf{y}}_{Wk}^H. \tag{4.24}$$

Now, each term in the sum in (4.24) is the criterion function for a single channel ML frequency estimator, since each vector $\breve{\mathbf{y}}_{Wk}$ contains a sum of cisoids in temporally white noise. The asymptotic properties of this estimator has been studied in [RZ93], where it was shown that it is strongly consistent. Hence, the minimizing argument of each term in the sum in (4.24) tends to $\boldsymbol{\omega}_0$ w.p.1 as $N \to \infty$. It then follows that the minimizing argument of the sum tends to $\boldsymbol{\omega}_0$ w.p.1 as $N \to \infty$. This proves consistency of the Doppler estimator in (4.13). Given that we can estimate $\boldsymbol{\omega}$ consistently it follows that

$$\begin{aligned} \hat{\boldsymbol{\Gamma}} &= \mathbf{Y}\mathbf{S}^H(\hat{\boldsymbol{\omega}}) \left[\mathbf{S}(\hat{\boldsymbol{\omega}})\mathbf{S}^H(\hat{\boldsymbol{\omega}})\right]^{-1} \\ &\to \mathbf{Y}\mathbf{S}^H(\boldsymbol{\omega}_0) \left[\mathbf{S}(\boldsymbol{\omega}_0)\mathbf{S}^H(\boldsymbol{\omega}_0)\right]^{-1} \to \boldsymbol{\Gamma} \end{aligned} \tag{4.25}$$

$$\begin{aligned} \hat{\mathbf{Q}} &= \frac{1}{N}(\mathbf{Y} - \hat{\boldsymbol{\Gamma}}\mathbf{S}(\hat{\boldsymbol{\omega}}))(\mathbf{Y} - \hat{\boldsymbol{\Gamma}}\mathbf{S}(\hat{\boldsymbol{\omega}}))^H \\ &\to \frac{1}{N}(\mathbf{Y} - \boldsymbol{\Gamma}\mathbf{S}(\boldsymbol{\omega}_0))(\mathbf{Y} - \boldsymbol{\Gamma}\mathbf{S}(\boldsymbol{\omega}_0))^H \to \mathbf{Q} \end{aligned} \tag{4.26}$$

w.p.1 as $N \to \infty$. Hence, all unknown parameters in (4.6) are consistently estimated.

Asymptotic efficiency

Since we have a consistent estimate of $\mathbf{Q}$ available we can write

$$\mathbf{W} = \mathbf{Q}^{-1} + \tilde{\mathbf{W}} \quad \text{with} \quad \|\tilde{\mathbf{W}}\| = o_p(1), \tag{4.27}$$

where $o_p(g_N)$ denotes a random variable that converges to zero in probability at a rate faster than $g_N \to 0$, see [Ful96]. The WLS criterion function can then be written as

$$\begin{aligned} J(\boldsymbol{\omega}) &= \mathrm{Tr}\{\mathbf{P}^{\perp}_{\mathbf{S}^H}\mathbf{Y}^H\mathbf{W}\mathbf{Y}\} \\ &= \mathrm{Tr}\{\mathbf{P}^{\perp}_{\mathbf{S}^H}\mathbf{Y}^H\mathbf{Q}^{-1}\mathbf{Y}\} + \mathrm{Tr}\Big\{\mathbf{P}^{\perp}_{\mathbf{S}^H}\mathbf{Y}^H\tilde{\mathbf{W}}\mathbf{Y}\Big\}. \end{aligned} \tag{4.28}$$

A Taylor expansion of the gradient of the criterion function shows that the asymptotic distribution of the estimates is governed by the behavior of the gradient and the Hessian as $N \to \infty$, see [Lju99, Section 9.2]. Since $\|\tilde{\mathbf{W}}\| = o_p(1)$, it is clear that the second term in (4.28) has a negligible (for large N) gradient and Hessian as compared to the first term. Thus, replacing $\mathbf{W}$ with $\mathbf{Q}$ will not change the asymptotic distribution of the Doppler estimates. Therefore, we can study the asymptotic distribution of

$$J(\boldsymbol{\omega}) = \mathrm{Tr}\{\mathbf{P}^{\perp}_{\mathbf{S}^H}\mathbf{Y}^H\mathbf{Q}^{-1}\mathbf{Y}\} = \mathrm{Tr}\{\mathbf{Y}_Q\mathbf{P}^{\perp}_{\mathbf{S}^H}\mathbf{Y}_Q^H\}, \tag{4.29}$$

where $\mathbf{Y}_Q = \mathbf{Q}^{-1/2}\mathbf{Y}$. Since the noise in $\mathbf{Y}_Q$ is spatially white, (4.29) is also the criterion function for the ML estimator.

Now, the model in (4.6) is a dual model of the standard DOA estimation model. To see this, take the transpose of (4.6)

$$\mathbf{Y}^T = \mathbf{S}^T(\boldsymbol{\omega})\boldsymbol{\Gamma}^T + \mathbf{N}^T \tag{4.30}$$

and compare it with the standard DOA estimation model

$$\mathbf{Y} = \mathbf{A}(\boldsymbol{\theta})\mathbf{X} + \mathbf{N}. \tag{4.31}$$

Hence, by replacing frequencies with DOAs and temporal samples with spatial samples we obtain the standard DOA estimation model. Therefore, studying the performance of ML frequency estimation as $N \to \infty$ is equivalent to the case of ML DOA estimation as $K \to \infty$. The latter case has been studied in [Vib95] under the assumption of spatio-temporally white noise. It was shown that the ML DOA estimator is asymptotically efficient as $K \to \infty$.

Since the noise in $\mathbf{Y}_Q$ is white and the uniform temporal sampling satisfies the assumption in [Vib95], it follows that the Doppler estimator considered here is asymptotically 1-D efficient as $N \to \infty$. It was shown in the previous chapter that the 1-D and 2-D Doppler CRBs are asymptotically equivalent as $N \to \infty$. It then follows that the Doppler estimator also is asymptotically 2-D efficient.

The 1-D DOA estimator

Consistency

Next, we prove consistency (in probability) for the DOA estimator in (4.20). The estimate of the k-th DOA can be expressed as

$$\begin{aligned}
\hat{\theta}_k(\hat{\boldsymbol{\omega}}) &= \arg\max_{\theta_k} \frac{\left|\mathbf{a}_k^H \mathbf{W} \hat{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}})\right|^2}{\mathbf{a}_k^H \mathbf{W} \mathbf{a}_k} \\
&= \arg\max_{\theta_k} \mathrm{Tr}\Big\{\mathbf{P}_{\mathbf{W}^{1/2}\mathbf{a}_k} \mathbf{W}^{1/2} \hat{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}}) \hat{\boldsymbol{\Gamma}}_k^H(\hat{\boldsymbol{\omega}}) \mathbf{W}^{1/2}\Big\} \\
&= \arg\min_{\theta_k} \mathrm{Tr}\Big\{\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_k} \mathbf{W}^{1/2} \hat{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}}) \hat{\boldsymbol{\Gamma}}_k^H(\hat{\boldsymbol{\omega}}) \mathbf{W}^{1/2}\Big\} \\
&= \arg\min_{\theta_k} V(\theta_k, \hat{\boldsymbol{\omega}}),
\end{aligned} \tag{4.32}$$

where

$$\begin{aligned}
\mathbf{a}_k &= \mathbf{a}(\theta_k) \\
\mathbf{P}_{\mathbf{W}^{1/2}\mathbf{a}_k} &= \mathbf{W}^{1/2}\mathbf{a}_k \mathbf{a}_k^H \mathbf{W}^{1/2} / (\mathbf{a}_k^H \mathbf{W} \mathbf{a}_k) \\
\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_k} &= \mathbf{I} - \mathbf{P}_{\mathbf{W}^{1/2}\mathbf{a}_k} \\
V(\theta_k, \hat{\boldsymbol{\omega}}) &= \mathrm{Tr}\Big\{\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_k} \mathbf{W}^{1/2} \hat{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}}) \hat{\boldsymbol{\Gamma}}_k^H(\hat{\boldsymbol{\omega}}) \mathbf{W}^{1/2}\Big\}.
\end{aligned} \tag{4.33}$$

In Appendix 4A, eq. (4A.30) it is shown that

$$\hat{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}}) = b_{0k}\mathbf{a}_{0k} + O_p(1/\sqrt{N}), \tag{4.34}$$

where $\mathbf{a}_{0k} \triangleq \mathbf{a}(\theta_{0k})$ and $O_p(g_N)$ denotes a random variable that converges to zero in probability at the same rate as $g_N \to 0$, see [Ful96]. This implies

$$\hat{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}}) \hat{\boldsymbol{\Gamma}}_k^H(\hat{\boldsymbol{\omega}}) = |b_{0k}|^2 \mathbf{a}_{0k}\mathbf{a}_{0k}^H + O_p(1/\sqrt{N}). \tag{4.35}$$

The criterion function $V(\theta_k, \hat{\boldsymbol{\omega}})$ converges in probability, uniformly in θ_k, to the limit function $\overline{V}(\theta_k)$, where

$$\overline{V}(\theta_k) = |b_{0k}|^2 \mathbf{a}_{0k}^H \mathbf{W}^{1/2} \mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_k} \mathbf{W}^{1/2} \mathbf{a}_{0k} \tag{4.36}$$

as $N \to \infty$. To show this, consider the difference

$$\begin{aligned}
&\sup_{\theta_k} \Big| \mathrm{Tr}\Big\{ \mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_k} \mathbf{W}^{1/2} \hat{\mathbf{\Gamma}}_k(\hat{\boldsymbol{\omega}}) \hat{\mathbf{\Gamma}}_k^H(\hat{\boldsymbol{\omega}}) \mathbf{W}^{1/2} \Big\} \\
&\quad - \mathrm{Tr}\Big\{ |b_{0k}|^2 \mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_k} \mathbf{W}^{1/2} \mathbf{a}_{0k} \mathbf{a}_{0k}^H \mathbf{W}^{1/2} \Big\} \Big| \\
&= \sup_{\theta_k} \Big| \mathrm{Tr}\Big\{ \mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_k} \mathbf{W}^{1/2} \left(\hat{\mathbf{\Gamma}}_k(\hat{\boldsymbol{\omega}}) \hat{\mathbf{\Gamma}}_k^H(\hat{\boldsymbol{\omega}}) - |b_{0k}|^2 \mathbf{a}_{0k} \mathbf{a}_{0k}^H \right) \mathbf{W}^{1/2} \Big\} \Big| \\
&\leq \sup_{\theta_k} \Big\| \mathbf{W}^{1/2} \mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_k} \mathbf{W}^{1/2} \Big\|_F \Big\| \hat{\mathbf{\Gamma}}_k(\hat{\boldsymbol{\omega}}) \hat{\mathbf{\Gamma}}_k^H(\hat{\boldsymbol{\omega}}) - |b_{0k}|^2 \mathbf{a}_{0k} \mathbf{a}_{0k}^H \Big\|_F . \quad (4.37)
\end{aligned}$$

Since $\|\mathbf{W}^{1/2} \mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_k} \mathbf{W}^{1/2}\|$ is bounded and $\hat{\mathbf{\Gamma}}_k(\hat{\boldsymbol{\omega}}) \hat{\mathbf{\Gamma}}_k^H(\hat{\boldsymbol{\omega}}) \to |b_{0k}|^2 \mathbf{a}_{0k} \mathbf{a}_{0k}^H$ in probability, the right hand side of (4.37) converges to zero in probability. Clearly, $\overline{V}(\theta_k)$ is minimized when $\mathbf{a}_k = c\, \mathbf{a}_{0k}$ for some scalar c. This implies that $\theta_k = \theta_{0k}$ provided that the array manifold is unambiguous. It then follows that $\hat{\theta}_k$ is a consistent estimate of θ_k (in probability).

Asymptotic efficiency

Next, we derive the asymptotic distribution of the 1-D DOA estimator in (4.20) and show that the variance of this distribution equals the asymptotic 2-D CRB. We compute the asymptotic distribution of the normalized DOA estimation error, $\sqrt{N}(\hat{\theta} - \theta_0)$, using standard Taylor expansion techniques, see e.g. [Lju99, Section 9.2]. Since $V'(\hat{\theta}, \hat{\boldsymbol{\omega}}) = 0$, we have

$$\sqrt{N}\, V'(\theta_0, \hat{\boldsymbol{\omega}}) + V''(\theta_0, \hat{\boldsymbol{\omega}}) \sqrt{N}(\hat{\theta} - \theta_0) + \frac{1}{2} V'''(\theta_*, \hat{\boldsymbol{\omega}}) \sqrt{N}(\hat{\theta} - \theta_0)^2 = 0 \quad (4.38)$$

for some θ_* between θ_0 and $\hat{\theta}$. The k subscript has been dropped for brevity. In Appendix 4A it is shown that the derivatives in (4.38) are of the following orders

$$\begin{aligned}
V'(\theta_0, \hat{\boldsymbol{\omega}}) &= O_p(1/\sqrt{N}) && (4.39) \\
V''(\theta_0, \hat{\boldsymbol{\omega}}) &= O_p(1) && (4.40) \\
V'''(\theta_*, \hat{\boldsymbol{\omega}}) &= O_p(1). && (4.41)
\end{aligned}$$

Since $\sqrt{N}\, V'(\theta_0, \hat{\boldsymbol{\omega}})$ and $V''(\theta_0, \hat{\boldsymbol{\omega}})$ are bounded, (4.38) implies that $\sqrt{N}(\hat{\theta} - \theta_0)$ must be bounded, i.e.

$$\hat{\theta} - \theta_0 = O_p(1/\sqrt{N}), \quad (4.42)$$

which in turn means that the third term in (4.38) can be neglected in the asymptotic analysis. Next, we show that replacing $\hat{\boldsymbol{\omega}}$ with $\boldsymbol{\omega}_0$ does not change

the asymptotic distribution of $\sqrt{N}(\hat{\theta}-\theta_0)$. In Appendix 4A it is shown that

$$\begin{aligned} V'(\theta_0,\hat{\boldsymbol{\omega}}) &= V'(\theta_0,\boldsymbol{\omega}_0)+o_p(1/\sqrt{N}) && (4.43)\\ V''(\theta_0,\hat{\boldsymbol{\omega}}) &= V''(\theta_0,\boldsymbol{\omega}_0)+o_p(1). && (4.44)\end{aligned}$$

Inserting this into (4.38) yields

$$\sqrt{N}V'(\theta_0,\boldsymbol{\omega}_0)+o_p(1)+\left[V''(\theta_0,\boldsymbol{\omega}_0)+o_p(1)\right]\sqrt{N}(\hat{\theta}-\theta_0)=0. \quad (4.45)$$

It is easily seen from this expression that the terms associated with replacing $\hat{\boldsymbol{\omega}}$ with $\boldsymbol{\omega}_0$ can be neglected in the asymptotic analysis. For large N we thus have

$$\sqrt{N}(\hat{\theta}-\theta_0)=-\frac{\sqrt{N}V'(\theta_0,\boldsymbol{\omega}_0)}{V''(\theta_0,\boldsymbol{\omega}_0)}+o_p(1)=-\frac{\sqrt{N}V'(\theta_0,\boldsymbol{\omega}_0)}{H}+o_p(1), \quad (4.46)$$

where

$$H=\lim_{N\to\infty}V''(\theta_0,\boldsymbol{\omega}_0) \quad (4.47)$$

in probability. From (4A.40) in Appendix 4A it follows that $\sqrt{N}V'(\theta_{0k},\boldsymbol{\omega}_0)$ is asymptotically normal, and hence from (4.46) that $\sqrt{N}(\hat{\theta}-\theta_0)$ is asymptotically normal. The limiting distribution is thus given by

$$\sqrt{N}(\hat{\theta}-\theta_0)\in \mathrm{As}\mathcal{N}(0,Q/H^2), \quad (4.48)$$

where

$$Q=\lim_{N\to\infty}N\mathrm{E}\left[(V'(\theta_0,\boldsymbol{\omega}_0))^2\right]. \quad (4.49)$$

In Appendix 4A it is shown that

$$Q=H=2|b_0|^2\mathbf{d}_0^H\mathbf{Q}^{-1/2}\mathbf{P}^{\perp}_{\mathbf{Q}^{-1/2}\mathbf{a}_0}\mathbf{Q}^{-1/2}\mathbf{d}_0 \quad (4.50)$$

so that the variance of the limiting distribution of $\sqrt{N}(\hat{\theta}-\theta_0)$ is given by

$$\frac{Q}{H^2}=\frac{1}{2|b_0|^2\mathbf{d}_0^H\mathbf{Q}^{-1/2}\mathbf{P}^{\perp}_{\mathbf{Q}^{-1/2}\mathbf{a}_0}\mathbf{Q}^{-1/2}\mathbf{d}_0}. \quad (4.51)$$

Except for the normalization factor $1/N$, this expressions equals the asymptotic 2-D CRB given in Theorem 3.4, since $\mathbf{s}_N^H\mathbf{s}_N=N$ when the sampling is uniform.

Strictly speaking, we have not shown that the limiting variance of the estimation error equals the CRB, only that the variance of the limiting distribution equals the CRB. This issue is more subtle and the interested reader is referred to [Ben93] and [Lju99, Appendix 9B] for a detailed discussion. □

Perhaps the most unexpected result is that the DOA parameter estimates are also asymptotically efficient, although the spatial parameterization was relaxed in the first step. The reason is that the "Doppler filtering" involved in forming $\hat{\mathbf{\Gamma}}$ effectively separates the contribution from each source, for large enough N, which results in efficiency of the simple beamformer (4.20). Although the results presented above are applicable for large N and exploiting the temporal dimension first, a similar method that starts by estimating $\boldsymbol{\theta}$ is simply derived using symmetry. The resulting method will be asymptotically efficient for large K.

4A Asymptotic Analysis of the 1-D/1-D Estimator

In this appendix, the details required for the proof of Theorem 4.5 are given. Asymptotic expressions for the criterion function $V(\theta_k, \hat{\boldsymbol{\omega}})$ and its derivatives w.r.t. θ_k are computed. The appendix is divided into two parts. In the first part, the asymptotic variance of the gradient, Q in (4.49), is computed. In the second part, the asymptotic Hessian, H in (4.47), is computed.

In the asymptotic analysis we will need the following results, which are formulated as a Lemma.

Lemma 4.1

$$\sum_{t=1}^{N} t^q e^{-j\omega t} = \begin{cases} O(N^q) & \omega \neq 0 \\ O(N^{q+1}) & \omega = 0 \end{cases} \tag{4A.1}$$

$$\sum_{t=1}^{N} \mathbf{n}(t) t^q e^{-j\omega t} = O_p(N^{q+1/2}) \quad \forall\, \omega \tag{4A.2}$$

$$\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0 = O_p(1/N\sqrt{N}) \tag{4A.3}$$

where $\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0$ is the Doppler frequency estimation error of the estimator in (4.13).

Proof To show (4A.1), we use the fact that for $\omega \neq 0$

$$\sum_{t=1}^{N} e^{-j\omega t} = \frac{1 - e^{-jN\omega}}{e^{j\omega} - 1}. \tag{4A.4}$$

Repeated differentiation w.r.t. ω on both sides in this expression then shows (4A.1) for $\omega \neq 0$. For $\omega = 0$ we use the following formula on p. 175 in [RW90]

$$\sum_{t=1}^{N} t^q = \frac{N^{q+1}}{q+1} + \frac{N^q}{2} + \ldots = O(N^{q+1}) \tag{4A.5}$$

and (4A.1) follows.

To show (4A.2), define

$$S_N = \sum_{t=1}^{N} n_k(t) t^q e^{-j\omega t}, \tag{4A.6}$$

where $n_k(t)$ is the k-th element in $\mathbf{n}(t)$, and compute the variance of S_N (note that S_N is zero mean)

$$\begin{aligned}\mathrm{E}\left[S_N S_N^*\right] &= \mathrm{E}\left[\sum_{t=1}^{N}\sum_{u=1}^{N} n_k(t)n_k^*(u)t^q u^q e^{-j\omega(t-u)}\right] = \sum_{t=1}^{N}\mathrm{E}\left[n_k(t)n_k^*(t)\right]t^{2q} \\ &= \sigma^2\sum_{t=1}^{N} t^{2q} = O(N^{2q+1}) \qquad \text{(4A.7)}\end{aligned}$$

using that the noise is temporally white and (4A.1). Applying Chebyshev's inequality (see e.g. [Ful96, Section 5.1]) then gives that $S_N = O_p(N^{q+1/2})$, and (4A.2) follows.

(4A.3) follows from Lemma 3.1 in [RZ93] and (4A.2). □

Calculation of Q

Differentiating the criterion function in (4.33) w.r.t. θ_k gives

$$V'(\theta_k, \hat{\boldsymbol{\omega}}) = \frac{\partial V(\theta_k, \hat{\boldsymbol{\omega}})}{\partial \theta_k} = \mathrm{Tr}\left\{\frac{\partial \mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_k}}{\partial \theta_k}\mathbf{W}^{1/2}\hat{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}})\hat{\boldsymbol{\Gamma}}_k^H(\hat{\boldsymbol{\omega}})\mathbf{W}^{1/2}\right\}. \qquad \text{(4A.8)}$$

Differentiating the projection matrix (see e.g. [VO91]) yields

$$\frac{\partial \mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_k}}{\partial \theta_k} = -\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_k}\mathbf{W}^{1/2}\mathbf{d}_k(\mathbf{a}_k^H\mathbf{W}\mathbf{a}_k)^{-1}\mathbf{a}_k^H\mathbf{W}^{1/2} - (\ldots)^H, \qquad \text{(4A.9)}$$

where the notation $(\ldots)^H$ means that the same expression appears again with complex conjugate transpose. Substituting into (4A.8) gives

$$V'(\theta_k, \hat{\boldsymbol{\omega}}) = -2\mathrm{Re}\left\{(\mathbf{a}_k^H\mathbf{W}\mathbf{a}_k)^{-1}\mathbf{a}_k^H\mathbf{W}\hat{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}})\hat{\boldsymbol{\Gamma}}_k^H(\hat{\boldsymbol{\omega}})\mathbf{W}^{1/2}\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_k}\mathbf{W}^{1/2}\mathbf{d}_k\right\}. \qquad \text{(4A.10)}$$

Since $\boldsymbol{\Gamma}_k = b_{0k}\mathbf{a}_{0k}$ we have $\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}}\mathbf{W}^{1/2}\boldsymbol{\Gamma}_k = \mathbf{0}$. Defining $\widetilde{\boldsymbol{\Gamma}}_k = \hat{\boldsymbol{\Gamma}}_k - \boldsymbol{\Gamma}_k$ we get

$$\begin{aligned}\hat{\boldsymbol{\Gamma}}_k\hat{\boldsymbol{\Gamma}}_k^H\mathbf{W}^{1/2}\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}} &= (\boldsymbol{\Gamma}_k + \widetilde{\boldsymbol{\Gamma}}_k)\hat{\boldsymbol{\Gamma}}_k^H\mathbf{W}^{1/2}\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}} \\ &= \boldsymbol{\Gamma}_k\hat{\boldsymbol{\Gamma}}_k^H\mathbf{W}^{1/2}\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}} + \widetilde{\boldsymbol{\Gamma}}_k(\boldsymbol{\Gamma}_k + \widetilde{\boldsymbol{\Gamma}}_k)^H\mathbf{W}^{1/2}\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}} \\ &= \boldsymbol{\Gamma}_k\hat{\boldsymbol{\Gamma}}_k^H\mathbf{W}^{1/2}\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}} + \widetilde{\boldsymbol{\Gamma}}_k\widetilde{\boldsymbol{\Gamma}}_k^H\mathbf{W}^{1/2}\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}} \\ &= \boldsymbol{\Gamma}_k\hat{\boldsymbol{\Gamma}}_k^H\mathbf{W}^{1/2}\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}} + o_p(|\widetilde{\boldsymbol{\Gamma}}_k|). \qquad \text{(4A.11)}\end{aligned}$$

Thus,

$$\begin{aligned} V'(\theta_{0k},\hat{\boldsymbol{\omega}}) &= -2\mathrm{Re}\Big\{(\mathbf{a}_{0k}^H\mathbf{W}\mathbf{a}_{0k})^{-1}\mathbf{a}_{0k}^H\mathbf{W}\boldsymbol{\Gamma}_k\hat{\boldsymbol{\Gamma}}_k^H(\hat{\boldsymbol{\omega}})\mathbf{W}^{1/2}\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}}\mathbf{W}^{1/2}\mathbf{d}_{0k}\Big\} \\ &+ o_p(|\widetilde{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}})|) \\ &= -2\mathrm{Re}\Big\{b_{0k}\hat{\boldsymbol{\Gamma}}_k^H(\hat{\boldsymbol{\omega}})\mathbf{W}^{1/2}\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}}\mathbf{W}^{1/2}\mathbf{d}_{0k}\Big\} + o_p(|\widetilde{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}})|) \\ &= -2\mathrm{Re}\Big\{b_{0k}\mathbf{d}_{0k}^H\mathbf{W}^{1/2}\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}}\mathbf{W}^{1/2}\hat{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}})\Big\} + o_p(|\widetilde{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}})|). \end{aligned} \tag{4A.12}$$

Next, in order to find a simple asymptotic expression for $\hat{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}})$, we analyze the asymptotic behavior of

$$\hat{\boldsymbol{\Gamma}}(\hat{\boldsymbol{\omega}}) = \mathbf{Y}\mathbf{S}^H(\hat{\boldsymbol{\omega}})\left[\mathbf{S}(\hat{\boldsymbol{\omega}})\mathbf{S}^H(\hat{\boldsymbol{\omega}})\right]^{-1}. \tag{4A.13}$$

The (m,n)-th element s_{mn} in the matrix $\mathbf{S}(\hat{\boldsymbol{\omega}})\mathbf{S}^H(\hat{\boldsymbol{\omega}})$ is given by

$$s_{mn} = \sum_{t=1}^{N} e^{j(\hat{\omega}_m-\hat{\omega}_n)t} = \begin{cases} N & m=n \\ -\dfrac{e^{jN(\hat{\omega}_m-\hat{\omega}_n)}-1}{e^{-j(\hat{\omega}_m-\hat{\omega}_n)}-1} & m\neq n. \end{cases} \tag{4A.14}$$

Decompose $\mathbf{S}(\hat{\boldsymbol{\omega}})\mathbf{S}^H(\hat{\boldsymbol{\omega}})$ into a sum of the diagonal part and the off-diagonal part according to

$$\mathbf{S}(\hat{\boldsymbol{\omega}})\mathbf{S}^H(\hat{\boldsymbol{\omega}}) = N\mathbf{I} + \mathbf{R}_N(\hat{\boldsymbol{\omega}}), \tag{4A.15}$$

where

$$\mathbf{R}_N(\hat{\boldsymbol{\omega}}) = \begin{bmatrix} 0 & s_{12} & \cdots & s_{1M} \\ s_{21} & 0 & & \vdots \\ \vdots & & \ddots & \vdots \\ s_{M1} & \cdots & \cdots & 0 \end{bmatrix} \tag{4A.16}$$

so that

$$\left[\mathbf{S}(\hat{\boldsymbol{\omega}})\mathbf{S}^H(\hat{\boldsymbol{\omega}})\right]^{-1} = \frac{1}{N}\left(\mathbf{I}+\frac{1}{N}\mathbf{R}_N(\hat{\boldsymbol{\omega}})\right)^{-1}. \tag{4A.17}$$

It easy to show that

$$\left(\mathbf{I}+\frac{1}{N}\mathbf{R}_N(\hat{\boldsymbol{\omega}})\right)^{-1} = \mathbf{I}-\frac{1}{N}\mathbf{R}_N(\hat{\boldsymbol{\omega}})+\frac{1}{N^2}\mathbf{R}_N(\hat{\boldsymbol{\omega}})^2-\ldots \tag{4A.18}$$

by the following multiplication

$$\left(\mathbf{I}+\frac{1}{N}\mathbf{R}_N\right)\left(\mathbf{I}-\frac{1}{N}\mathbf{R}_N+\frac{1}{N^2}\mathbf{R}_N^2-\ldots\right)=$$
$$\mathbf{I}-\frac{1}{N}\mathbf{R}_N+\frac{1}{N}\mathbf{R}_N+\frac{1}{N^2}\mathbf{R}_N^2-\frac{1}{N^2}\mathbf{R}_N^2-\ldots=\mathbf{I} \tag{4A.19}$$

provided $\|\frac{1}{N}\mathbf{R}_N\|$ is sufficiently small for the sum to converge. We thus have

$$\hat{\mathbf{\Gamma}}(\hat{\boldsymbol{\omega}})=\frac{1}{N}\mathbf{Y}\mathbf{S}^H(\hat{\boldsymbol{\omega}})[1+O_p(1/N)], \tag{4A.20}$$

since $\|\mathbf{R}_N(\hat{\boldsymbol{\omega}})\|$ is of order $O_p(1)$ because the Doppler frequencies are assumed to be distinct. Define $\mathbf{g}(\hat{\omega}_k)$ to be the k-th column of $\frac{1}{N}\mathbf{Y}\mathbf{S}^H(\hat{\boldsymbol{\omega}})$, i.e.

$$\mathbf{g}(\hat{\omega}_k)=\frac{1}{N}\sum_{t=1}^{N}\mathbf{y}(t)e^{-j\hat{\omega}_k t}. \tag{4A.21}$$

It then follows that

$$\hat{\mathbf{\Gamma}}_k(\hat{\boldsymbol{\omega}})=\mathbf{g}(\hat{\omega}_k)\left[1+O_p(1/N)\right]. \tag{4A.22}$$

Recalling that

$$\mathbf{y}(t) \;=\; \mathbf{A}(\boldsymbol{\theta}_0)\mathbf{B}_0\mathbf{s}(\boldsymbol{\omega}_0,t)+\mathbf{n}(t)=\sum_{m=1}^{M}b_{0m}\mathbf{a}_{0m}e^{j\omega_{0m}t}+\mathbf{n}(t), \tag{4A.23}$$

we obtain

$$\begin{aligned}\mathbf{g}(\hat{\omega}_k) &= \frac{1}{N}\sum_{t=1}^{N}\left(\sum_{m=1}^{M}b_{0m}\mathbf{a}_{0m}e^{j\omega_{0m}t}+\mathbf{n}(t)\right)e^{-j\hat{\omega}_k t}\\ &= b_{0k}\mathbf{a}_{0k}\frac{1}{N}\sum_{t=1}^{N}e^{j(\omega_{0k}-\hat{\omega}_k)t}+\sum_{\substack{m=1\\m\neq k}}^{M}b_{0m}\mathbf{a}_{0m}\frac{1}{N}\sum_{t=1}^{N}e^{j(\omega_{0m}-\hat{\omega}_k)t}+\\ &\quad \frac{1}{N}\sum_{t=1}^{N}\mathbf{n}(t)e^{-j\hat{\omega}_k t}=\mathbf{g}_1+\mathbf{g}_2+\mathbf{g}_3.\end{aligned} \tag{4A.24}$$

We have divided $\mathbf{g}(\hat{\omega}_k)$ into three parts: $\mathbf{g}_1$ which stems from the signal of interest, $\mathbf{g}_2$ which accounts for all other signals, and $\mathbf{g}_3$ which represents the contribution from the noise. The asymptotic behavior of these different terms is now analyzed using the Taylor expansion

$$e^{-j\hat{\omega}_k t}=e^{-j\omega_{0k}t}-j(\hat{\omega}_k-\omega_{0k})te^{-j\omega_{0k}t}+\ldots. \tag{4A.25}$$

From Lemma 4.1 it follows that each additional higher order term in this expansion contributes with an additional factor $t(\hat{\omega}_k - \omega_{0k})$ under the $\sum_t$ sum. This will decrease the order of the sum by $O(N)\, O_p(1/N\sqrt{N}) = O_p(1/\sqrt{N})$. These higher order terms can therefore be neglected in the asymptotic analysis. Consequently,

$$\mathbf{g}_1 = b_{0k}\mathbf{a}_{0k}\frac{1}{N}\left(N - j\underbrace{(\hat{\omega}_k - \omega_{0k})}_{O_p\left(\frac{1}{N\sqrt{N}}\right)}\underbrace{\sum_{t=1}^{N} t}_{O(N^2)} + \ldots \right) = b_{0k}\mathbf{a}_{0k}\left(1 + O_p(1/\sqrt{N})\right) \tag{4A.26}$$

$$\begin{aligned}\mathbf{g}_2 &= \sum_{\substack{m=1\\ m\neq k}}^{M} b_{0m}\mathbf{a}_{0m}\frac{1}{N}\left(\underbrace{\sum_{t=1}^{N} e^{j(\omega_{0m}-\omega_{0k})t}}_{O(1)} - j\underbrace{(\hat{\omega}_k - \omega_{0k})}_{O_p\left(\frac{1}{N\sqrt{N}}\right)}\underbrace{\sum_{t=1}^{N} te^{j(\omega_{0m}-\omega_{0k})t}}_{O(N)} + \ldots \right) \\ &= O_p(1/N)\end{aligned} \tag{4A.27}$$

$$\begin{aligned}\mathbf{g}_3 &= \frac{1}{N}\left(\underbrace{\sum_{t=1}^{N} \mathbf{n}(t)e^{-j\omega_{0k}t}}_{O_p(\sqrt{N})} - j\underbrace{(\hat{\omega}_k - \omega_{0k})}_{O_p\left(\frac{1}{N\sqrt{N}}\right)}\underbrace{\sum_{t=1}^{N} \mathbf{n}(t)te^{-j\omega_{0k}t}}_{O_p(N\sqrt{N})} + \ldots \right) \\ &= \frac{1}{N}\sum_{t=1}^{N} \mathbf{n}(t)e^{-j\omega_{0k}t} + O_p(1/N).\end{aligned} \tag{4A.28}$$

Substituting these expression into (4A.24) yields

$$\mathbf{g}(\hat{\omega}_k) = b_{0k}\mathbf{a}_{0k}\left(1 + O_p(1/\sqrt{N})\right) + \frac{1}{N}\sum_{t=1}^{N} \mathbf{n}(t)e^{-j\omega_{0k}t} + O_p(1/N). \tag{4A.29}$$

These results are now used for computing asymptotic expressions for $\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}}\mathbf{W}^{1/2}\hat{\mathbf{\Gamma}}_k(\hat{\boldsymbol{\omega}})$ and $\widetilde{\mathbf{\Gamma}}_k(\hat{\boldsymbol{\omega}})$ which are needed for computing an asymp-

totic expression for $V'(\theta_{0k}, \hat{\boldsymbol{\omega}})$. Combining (4A.22) and (4A.29) we obtain

$$\begin{aligned}\hat{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}}) &= \left[b_{0k}\mathbf{a}_{0k}\left(1+O_p(1/\sqrt{N})\right)+\frac{1}{N}\sum_{t=1}^{N}\mathbf{n}(t)e^{-j\omega_{0k}t}+O_p(1/N)\right]\\ &\times\ [1+O_p(1/N)]\\ &= b_{0k}\mathbf{a}_{0k}\left(1+O_p(1/\sqrt{N})\right)+\frac{1}{N}\sum_{t=1}^{N}\mathbf{n}(t)e^{-j\omega_{0k}t}+O_p(1/N)\end{aligned} \tag{4A.30}$$

and, since $\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}}\mathbf{W}^{1/2}\mathbf{a}_{0k}=\mathbf{0}$,

$$\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}}\mathbf{W}^{1/2}\hat{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}})=\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}}\mathbf{W}^{1/2}\frac{1}{N}\sum_{t=1}^{N}\mathbf{n}(t)e^{-j\omega_{0k}t}+O_p(1/N). \tag{4A.31}$$

Using (4A.2), (4A.30) and the definition of $\widetilde{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}})$ we obtain

$$\widetilde{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}}) = \hat{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}})-b_{0k}\mathbf{a}_{0k}=O_p(1/\sqrt{N}). \tag{4A.32}$$

Substituting (4A.31) and (4A.32) into (4A.12) gives

$$\begin{aligned}V'(\theta_{0k},\hat{\boldsymbol{\omega}}) &= -2\mathrm{Re}\left\{b_{0k}\mathbf{d}_{0k}^{H}\mathbf{W}^{1/2}\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}}\mathbf{W}^{1/2}\frac{1}{N}\sum_{t=1}^{N}\mathbf{n}(t)e^{-j\omega_{0k}t}\right\}\\ &+\ o_p(1/\sqrt{N}).\end{aligned} \tag{4A.33}$$

Now, since

$$\mathbf{W}=\mathbf{Q}^{-1}+o_p(1)\quad\text{and}\quad\frac{1}{N}\sum_{t=1}^{N}\mathbf{n}(t)e^{-j\omega_{0k}t}=O_p(1/\sqrt{N}) \tag{4A.34}$$

we obtain

$$\begin{aligned}V'(\theta_{0k},\hat{\boldsymbol{\omega}}) &= -2\mathrm{Re}\left\{b_{0k}\mathbf{d}_{0k}^{H}\mathbf{Q}^{-1/2}\mathbf{P}^{\perp}_{\mathbf{Q}^{-1/2}\mathbf{a}_{0k}}\mathbf{Q}^{-1/2}\frac{1}{N}\sum_{t=1}^{N}\mathbf{n}(t)e^{-j\omega_{0k}t}\right\}\\ &+\ o_p(1/\sqrt{N}).\end{aligned} \tag{4A.35}$$

It can be seen from this expression that $V'(\theta_{0k},\hat{\boldsymbol{\omega}})$ is of order $O_p(1/\sqrt{N})$ which proves (4.39). To show (4.43), we compare (4A.35) with

$$V'(\theta_{0k},\boldsymbol{\omega}_0)=-2\mathrm{Re}\left\{b_{0k}\mathbf{d}_{0k}^{H}\mathbf{W}^{1/2}\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}}\mathbf{W}^{1/2}\hat{\boldsymbol{\Gamma}}_k(\boldsymbol{\omega}_0)\right\}+o_p(|\widetilde{\boldsymbol{\Gamma}}_k(\boldsymbol{\omega}_0)|). \tag{4A.36}$$

Using

$$\begin{aligned}\hat{\mathbf{\Gamma}}(\boldsymbol{\omega}_0) &= \mathbf{Y}\mathbf{S}^H(\boldsymbol{\omega}_0)\left[\mathbf{S}(\boldsymbol{\omega}_0)\mathbf{S}^H(\boldsymbol{\omega}_0)\right]^{-1} \\ &= \left[\mathbf{A}(\boldsymbol{\theta}_0)\mathbf{B}_0\mathbf{S}(\boldsymbol{\omega}_0)+\mathbf{E}\right]\mathbf{S}^H(\boldsymbol{\omega}_0)\left[\mathbf{S}(\boldsymbol{\omega}_0)\mathbf{S}^H(\boldsymbol{\omega}_0)\right]^{-1} \\ &= \mathbf{A}(\boldsymbol{\theta}_0)\mathbf{B}_0+\mathbf{E}\mathbf{S}^H(\boldsymbol{\omega}_0)\left[\mathbf{S}(\boldsymbol{\omega}_0)\mathbf{S}^H(\boldsymbol{\omega}_0)\right]^{-1} \\ &= \mathbf{A}(\boldsymbol{\theta}_0)\mathbf{B}_0+\frac{1}{N}\mathbf{E}\mathbf{S}^H(\boldsymbol{\omega}_0)\left[1+O(1/N)\right], \qquad (4A.37)\end{aligned}$$

we have that

$$\begin{aligned}\hat{\mathbf{\Gamma}}_k(\boldsymbol{\omega}_0) &= b_{0k}\mathbf{a}_{0k}+\frac{1}{N}\sum_{t=1}^{N}\mathbf{n}(t)e^{-j\omega_{0k}t}\left[1+O(1/N)\right] \\ &= b_{0k}\mathbf{a}_{0k}+\frac{1}{N}\sum_{t=1}^{N}\mathbf{n}(t)e^{-j\omega_{0k}t}+O_p(1/N\sqrt{N}), \qquad (4A.38)\end{aligned}$$

and

$$\begin{aligned}\widetilde{\mathbf{\Gamma}}_k(\boldsymbol{\omega}_0) &= \hat{\mathbf{\Gamma}}_k(\boldsymbol{\omega}_0)-\mathbf{\Gamma}_k \\ &= b_{0k}\mathbf{a}_{0k}+\frac{1}{N}\sum_{t=1}^{N}\mathbf{n}(t)e^{-j\omega_{0k}t}\left[1+O(1/N)\right]-b_{0k}\mathbf{a}_{0k} \\ &= O_p(1/\sqrt{N}). \qquad (4A.39)\end{aligned}$$

Substituting (4A.38) and (4A.39) into (4A.36) gives

$$\begin{aligned}V'(\theta_{0k},\boldsymbol{\omega}_0) &= -2\mathrm{Re}\left\{b_{0k}\mathbf{d}_{0k}^H\mathbf{W}^{1/2}\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_{0k}}\mathbf{W}^{1/2}\frac{1}{N}\sum_{t=1}^{N}\mathbf{n}(t)e^{-j\omega_{0k}t}\right\} \\ &+ o_p(1/\sqrt{N}) \\ &= -2\mathrm{Re}\left\{b_{0k}\mathbf{d}_{0k}^H\mathbf{Q}^{-1/2}\mathbf{P}^{\perp}_{\mathbf{Q}^{-1/2}\mathbf{a}_{0k}}\mathbf{Q}^{-1/2}\frac{1}{N}\sum_{t=1}^{N}\mathbf{n}(t)e^{-j\omega_{0k}t}\right\} \\ &+ o_p(1/\sqrt{N}). \qquad (4A.40)\end{aligned}$$

Comparing (4A.35) and (4A.40) we see that

$$V'(\theta_{0k},\hat{\boldsymbol{\omega}}) = V'(\theta_{0k},\boldsymbol{\omega}_0)+o_p(1/\sqrt{N}), \qquad (4A.41)$$

which proves (4.43).

We are now ready to compute the final expression for Q. Since the first term in (4A.40) is of order $O_p(1/\sqrt{N})$, the $o_p(1/\sqrt{N})$ term can be ignored

when computing Q. Thus, for sufficiently large N we have

$$\begin{aligned}
&N\mathrm{E}\left[(V'(\theta_{0k},\boldsymbol{\omega}_0))^2\right] = \\
&= \frac{4}{N}\mathrm{E}\Bigg[\mathrm{Re}\Big\{b_{0k}\mathbf{d}_{0k}^H\mathbf{Q}^{-1/2}\mathbf{P}^{\perp}_{\mathbf{Q}^{-1/2}\mathbf{a}_{0k}}\mathbf{Q}^{-1/2}\sum_{t=1}^{N}\mathbf{n}(t)e^{-j\omega_{0k}t}\Big\} \\
&\times \mathrm{Re}\Big\{b_{0k}\mathbf{d}_{0k}^H\mathbf{Q}^{-1/2}\mathbf{P}^{\perp}_{\mathbf{Q}^{-1/2}\mathbf{a}_{0k}}\mathbf{Q}^{-1/2}\sum_{u=1}^{N}\mathbf{n}(u)e^{-j\omega_{0k}u}\Big\}\Bigg] \\
&= \frac{2}{N}\mathrm{E}\Bigg[\mathrm{Re}\Big\{|b_{0k}|^2\mathbf{d}_{0k}^H\mathbf{Q}^{-1/2}\mathbf{P}^{\perp}_{\mathbf{Q}^{-1/2}\mathbf{a}_{0k}}\mathbf{Q}^{-1/2}\sum_{t=1}^{N}\mathbf{n}(t)e^{-j\omega_{0k}t} \\
&\times \sum_{u=1}^{N}\mathbf{n}^H(u)e^{j\omega_{0k}u}\mathbf{Q}^{-1/2}\mathbf{P}^{\perp}_{\mathbf{Q}^{-1/2}\mathbf{a}_{0k}}\mathbf{Q}^{-1/2}\mathbf{d}_{0k}\Big\}\Bigg] \\
&+ \frac{2}{N}\mathrm{E}\Bigg[\mathrm{Re}\Big\{b_{0k}^2\mathbf{d}_{0k}^H\mathbf{Q}^{-1/2}\mathbf{P}^{\perp}_{\mathbf{Q}^{-1/2}\mathbf{a}_{0k}}\mathbf{Q}^{-1/2}\sum_{t=1}^{N}\mathbf{n}(t)e^{-j\omega_{0k}t} \\
&\times \sum_{u=1}^{N}\mathbf{n}^T(u)e^{-j\omega_{0k}u}\mathbf{Q}^{-T/2}\mathbf{P}^{\perp T}_{\mathbf{Q}^{-1/2}\mathbf{a}_{0k}}\mathbf{Q}^{-T/2}\mathbf{d}_{0k}^*\Big\}\Bigg]. \qquad (4A.42)
\end{aligned}$$

Since,

$$\sum_{t=1}^{N}\sum_{u=1}^{N}\mathrm{E}[\mathbf{n}(t)\mathbf{n}^H(u)]e^{j\omega_{0k}(u-t)} = \sum_{t=1}^{N}\mathrm{E}[\mathbf{n}(t)\mathbf{n}^H(t)] = N\mathbf{Q} \qquad (4A.43)$$

and $\mathrm{E}[\mathbf{n}(t)\mathbf{n}^T(u)] = \mathbf{0}$, we obtain the final expression for Q according to

$$Q = \lim_{N\to\infty} N\ \mathrm{E}\left[(V'(\theta_{0k},\boldsymbol{\omega}_0))^2\right] = 2|b_{0k}|^2\mathbf{d}_{0k}^H\mathbf{Q}^{-1/2}\mathbf{P}^{\perp}_{\mathbf{Q}^{-1/2}\mathbf{a}_{0k}}\mathbf{Q}^{-1/2}\mathbf{d}_{0k}. \qquad (4A.44)$$

This proves Q in (4.50).

Calculation of H

Differentiating the criterion function in (4.33) w.r.t. θ_k twice gives

$$V''(\theta_k,\hat{\boldsymbol{\omega}}) = \mathrm{Tr}\left\{\frac{\partial^2\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_k}}{\partial\theta_k^2}\mathbf{W}^{1/2}\hat{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}})\hat{\boldsymbol{\Gamma}}_k^H(\hat{\boldsymbol{\omega}})\mathbf{W}^{1/2}\right\}. \qquad (4A.45)$$

From (4A.30) and (4A.38) we have that

$$\hat{\boldsymbol{\Gamma}}_k(\hat{\boldsymbol{\omega}}) = b_{0k}\mathbf{a}_{0k} + o_p(1) \qquad (4A.46)$$

$$\hat{\boldsymbol{\Gamma}}_k(\boldsymbol{\omega}_0) = b_{0k}\mathbf{a}_{0k} + o_p(1) \qquad (4A.47)$$

so that

$$\hat{\mathbf{\Gamma}}_k(\hat{\boldsymbol{\omega}}) = \hat{\mathbf{\Gamma}}_k(\boldsymbol{\omega}_0) + o_p(1) \tag{4A.48}$$

and

$$\hat{\mathbf{\Gamma}}_k(\hat{\boldsymbol{\omega}})\hat{\mathbf{\Gamma}}_k^H(\hat{\boldsymbol{\omega}}) = \hat{\mathbf{\Gamma}}_k(\boldsymbol{\omega}_0)\hat{\mathbf{\Gamma}}_k^H(\boldsymbol{\omega}_0) + o_p(1) \tag{4A.49}$$

which shows that

$$V''(\theta_k, \hat{\boldsymbol{\omega}}) = V''(\theta_k, \boldsymbol{\omega}_0) + o_p(1) \tag{4A.50}$$

and (4.44) follows. Defining $\mathbf{a}_W = \mathbf{W}^{1/2}\mathbf{a}_{0k}$ and $\mathbf{d}_W = \mathbf{W}^{1/2}\mathbf{d}_{0k}$ we obtain from [VO91]

$$\begin{aligned}\left.\frac{\partial^2 \mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_k}}{\partial\theta_k^2}\right|_{\theta_k=\theta_{0k}} &= \mathbf{P}^{\perp}_{\mathbf{a}_W}\mathbf{d}_W\mathbf{a}_W^{\dagger}\mathbf{d}_W\mathbf{a}_W^{\dagger} + \mathbf{a}_W^{\dagger H}\mathbf{d}_W^H\mathbf{P}^{\perp}_{\mathbf{a}_W}\mathbf{d}_W\mathbf{a}_W^{\dagger} \\ &-\mathbf{P}^{\perp}_{\mathbf{a}_W}\mathbf{d}_W^2\mathbf{a}_W^{\dagger} - \mathbf{P}^{\perp}_{\mathbf{a}_W}\mathbf{d}_W(\mathbf{a}_W^H\mathbf{a}_W)^{-1}\mathbf{d}_W^H\mathbf{P}^{\perp}_{\mathbf{a}_W} + \mathbf{P}^{\perp}_{\mathbf{a}_W}\mathbf{d}_W\mathbf{a}_W^{\dagger}\mathbf{d}_W\mathbf{a}_W^{\dagger} \\ &+(\ldots)^H,\end{aligned} \tag{4A.51}$$

where $\mathbf{a}_W^{\dagger} = (\mathbf{a}_W^H\mathbf{a}_W)^{-1}\mathbf{a}_W^H$ and

$$\mathbf{d}_W^2 = \mathbf{W}^{1/2}\left.\frac{\partial^2\mathbf{a}(\theta_k)}{\partial\theta_k^2}\right|_{\theta_k=\theta_{0k}}. \tag{4A.52}$$

If the derivatives up to second order of the array manifold are bounded, we conclude that $V''(\theta_{0k}, \boldsymbol{\omega}_0)$ is bounded since $\hat{\mathbf{\Gamma}}_k(\boldsymbol{\omega}_0)\hat{\mathbf{\Gamma}}_k^H(\boldsymbol{\omega}_0)$ is bounded. Similarly, we conclude that

$$V'''(\theta_{*k}, \boldsymbol{\omega}_0) = \mathrm{Tr}\left\{\left.\frac{\partial^3\mathbf{P}^{\perp}_{\mathbf{W}^{1/2}\mathbf{a}_k}}{\partial\theta_k^3}\right|_{\theta_k=\theta_{*k}}\mathbf{W}^{1/2}\hat{\mathbf{\Gamma}}_k(\boldsymbol{\omega}_0)\hat{\mathbf{\Gamma}}_k^H(\boldsymbol{\omega}_0)\mathbf{W}^{1/2}\right\} \tag{4A.53}$$

for some θ_{*k} between θ_{0k} and $\hat{\theta}_k$, is bounded if the derivatives up to third order of the array manifold are bounded. We have thus shown (4.40) and (4.41). From (4A.38) we have

$$\lim_{N\to\infty}\hat{\mathbf{\Gamma}}_k(\boldsymbol{\omega}_0) = b_{0k}\mathbf{a}_{0k} \tag{4A.54}$$

in probability. Hence,

$$\lim_{N\to\infty}\hat{\mathbf{\Gamma}}_k(\boldsymbol{\omega}_0)\hat{\mathbf{\Gamma}}_k^H(\boldsymbol{\omega}_0) = |b_{0k}|^2\mathbf{a}_{0k}\mathbf{a}_{0k}^H \tag{4A.55}$$

in probability. Defining $\mathbf{a}_Q = \mathbf{Q}^{-1/2}\mathbf{a}_{0k}$ and $\mathbf{d}_Q = \mathbf{Q}^{-1/2}\mathbf{d}_{0k}$ the terms in $H = \lim_{N\to\infty} V''(\theta_{0k}, \boldsymbol{\omega}_0)$ are given by

$$\begin{aligned}
&\lim_{N\to\infty} \mathrm{Tr}\Big\{\mathbf{P}^{\perp}_{\mathbf{a}_W}\mathbf{d}_W\mathbf{a}_W^{\dagger}\mathbf{d}_W\mathbf{a}_W^{\dagger}\mathbf{W}^{1/2}\hat{\mathbf{\Gamma}}_k(\boldsymbol{\omega}_0)\hat{\mathbf{\Gamma}}_k^H(\boldsymbol{\omega}_0)\mathbf{W}^{1/2}\Big\} = \\
&\quad = \mathrm{Tr}\Big\{|b_{0k}|^2\mathbf{P}^{\perp}_{\mathbf{a}_Q}\mathbf{d}_Q\mathbf{a}_Q^{\dagger}\mathbf{d}_Q\mathbf{a}_Q^H\Big\} = |b_{0k}|^2\mathbf{a}_Q^H\mathbf{P}^{\perp}_{\mathbf{a}_Q}\mathbf{d}_Q\mathbf{a}_Q^{\dagger}\mathbf{d}_Q\mathbf{a}_Q^H = 0 \\
&\lim_{N\to\infty} \mathrm{Tr}\Big\{\mathbf{a}_W^{\dagger H}\mathbf{d}_W^H\mathbf{P}^{\perp}_{\mathbf{a}_W}\mathbf{d}_W\mathbf{a}_W^{\dagger}\mathbf{W}^{1/2}\hat{\mathbf{\Gamma}}_k(\boldsymbol{\omega}_0)\hat{\mathbf{\Gamma}}_k^H(\boldsymbol{\omega}_0)\mathbf{W}^{1/2}\Big\} = \\
&\quad = \mathrm{Tr}\Big\{|b_{0k}|^2\mathbf{a}_Q^{\dagger H}\mathbf{d}_Q^H\mathbf{P}^{\perp}_{\mathbf{a}_Q}\mathbf{d}_Q\mathbf{a}_Q^H\Big\} = |b_{0k}|^2\mathbf{d}_Q^H\mathbf{P}^{\perp}_{\mathbf{a}_Q}\mathbf{d}_Q \\
&\lim_{N\to\infty} \mathrm{Tr}\Big\{\mathbf{P}^{\perp}_{\mathbf{a}_W}\mathbf{d}_W^2\mathbf{a}_W^{\dagger}\mathbf{W}^{1/2}\hat{\mathbf{\Gamma}}_k(\boldsymbol{\omega}_0)\hat{\mathbf{\Gamma}}_k^H(\boldsymbol{\omega}_0)\mathbf{W}^{1/2}\Big\} = \\
&\quad = \mathrm{Tr}\Big\{|b_{0k}|^2\mathbf{P}^{\perp}_{\mathbf{a}_Q}\mathbf{d}_Q^2\mathbf{a}_Q^H\Big\} = |b_{0k}|^2\mathbf{a}_Q^H\mathbf{P}^{\perp}_{\mathbf{a}_Q}\mathbf{d}_Q^2 = 0 \\
&\lim_{N\to\infty} \mathrm{Tr}\Big\{\mathbf{P}^{\perp}_{\mathbf{a}_W}\mathbf{d}_W(\mathbf{a}_W^H\mathbf{a}_W)^{-1}\mathbf{d}_W^H\mathbf{P}^{\perp}_{\mathbf{a}_W}\mathbf{W}^{1/2}\hat{\mathbf{\Gamma}}_k(\boldsymbol{\omega}_0)\hat{\mathbf{\Gamma}}_k^H(\boldsymbol{\omega}_0)\mathbf{W}^{1/2}\Big\} = \\
&\quad = \mathrm{Tr}\Big\{|b_{0k}|^2\mathbf{P}^{\perp}_{\mathbf{a}_Q}\mathbf{d}_Q(\mathbf{a}_Q^H\mathbf{a}_Q)^{-1}\mathbf{d}_Q^H\mathbf{P}^{\perp}_{\mathbf{a}_Q}\mathbf{a}_Q\mathbf{a}_Q^H\Big\} = 0 \\
&\lim_{N\to\infty} \mathrm{Tr}\Big\{\mathbf{P}^{\perp}_{\mathbf{a}_W}\mathbf{d}_W\mathbf{a}_W^{\dagger}\mathbf{d}_W\mathbf{a}_W^{\dagger}\mathbf{W}^{1/2}\hat{\mathbf{\Gamma}}_k(\boldsymbol{\omega}_0)\hat{\mathbf{\Gamma}}_k^H(\boldsymbol{\omega}_0)\mathbf{W}^{1/2}\Big\} = \\
&\quad = \mathrm{Tr}\Big\{|b_{0k}|^2\mathbf{P}^{\perp}_{\mathbf{a}_Q}\mathbf{d}_Q\mathbf{a}_Q^{\dagger}\mathbf{d}_Q\mathbf{a}_Q^H\Big\} = |b_{0k}|^2\mathbf{a}_Q^H\mathbf{P}^{\perp}_{\mathbf{a}_Q}\mathbf{d}_Q\mathbf{a}_Q^{\dagger}\mathbf{d}_Q = 0
\end{aligned}$$

so that

$$\begin{aligned}
H &= \lim_{N\to\infty} V''(\theta_{0k}, \boldsymbol{\omega}_0) = 2|b_{0k}|^2\mathbf{d}_Q^H\mathbf{P}^{\perp}_{\mathbf{a}_Q}\mathbf{d}_Q \\
&= 2|b_{0k}|^2\mathbf{d}_{0k}^H\mathbf{Q}^{-1/2}\mathbf{P}^{\perp}_{\mathbf{Q}^{-1/2}\mathbf{a}_{0k}}\mathbf{Q}^{-1/2}\mathbf{d}_{0k} \qquad (4\mathrm{A}.56)
\end{aligned}$$

which proves H in (4.50). □

Chapter 5

Computing the Estimates

The methods presented herein are all implicit in nature. The parameter estimates are defined as the minimizing arguments of some nonlinear criterion function. In practice, finding the minimum is far from trivial, especially when the number of parameters is large. The presented objective functions are all of the nonlinear least squares type. We will in the following suggest a computationally attractive and reasonably reliable procedure to solve such problems. The technique is based on local numerical optimization (e.g. Gauss-Newton [DS83]), initialized using the so-called RELAX procedure [LZS93]. In addition to solving the optimization problem, the number of targets, M, must also be determined in any practical situation. Methods for estimating M are proposed in, e.g., [Shu83, Böh95, CH96, Eri01, Eri02], and will not be further explored here.

5.1 Local Optimization

The 2-D ML criterion in (3.40) can be efficiently implemented by first performing a QR decomposition of $\boldsymbol{\Upsilon}$ according to

$$\boldsymbol{\Upsilon} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix}, \tag{5.1}$$

and then computing the ML criterion function as

$$V_{2D} \;=\; \left\| \mathbf{y}_N - \mathbf{Q}_1 \left(\mathbf{Q}_1^H \mathbf{y}_N \right) \right\|_F^2. \tag{5.2}$$

Since only $\mathbf{Q}_1$ is needed for computing the criterion function an "economy size" QR decomposition can be used. This least squares criterion can then be minimized using some numerical optimization method, e.g. Gauss-Newton.

Similarly, the 1-D Doppler problem in (4.13) can be implemented by first performing a QR decomposition of $\mathbf{S}^H$ according to

$$\mathbf{S}^H = \left[\begin{array}{cc} \mathbf{Q}_1 & \mathbf{Q}_2 \end{array} \right] \left[\begin{array}{c} \mathbf{R}_1 \\ \mathbf{0} \end{array} \right], \tag{5.3}$$

and then minimizing the least squares criterion

$$V_{\text{Doppler}} = \left\| \mathbf{Y} - (\mathbf{Y}\mathbf{Q}_1)\,\mathbf{Q}_1^H \right\|_F^2. \tag{5.4}$$

The subsequent DOA estimator is then implemented by minimizing

$$V_{\text{DOA}} = \left\| \hat{\mathbf{\Gamma}}_m - \mathbf{a}^H \hat{\mathbf{\Gamma}}_m \mathbf{a} / (\mathbf{a}^H \mathbf{a}) \right\|_F^2, \quad m = 1, \ldots, M. \tag{5.5}$$

For each m, this is done by a coarse grid search (FFT in the case of a ULA) followed by a local optimization. The computational complexity of the above nonlinear LS problems can be further reduced by explicitly exploiting the structure of separable least squares, see e.g. [GP73, Kau75, RW80].

5.2 Computing Initial Estimates

We will use the RELAX method to find initial estimates for the WLS estimator. RELAX is closely related to the Alternating Projection (AP) algorithm [ZW88]. The difference is that RELAX updates the parameters of only one target at the time, whereas AP updates the nonlinear (e.g. Doppler) parameter of one target and the linear ("amplitude") parameters of *all* targets at the same time.

We will only give the details for the 1-D problem (4.9). The extension to the 2-D case is obvious, but involves sequentially searching over two parameters (DOA and Doppler) for each target. To stress the dependence on the target parameters, (4.9) is expressed as

$$V(\mathbf{\Gamma}_W, \boldsymbol{\omega}) = \|\mathbf{Y}_W - \mathbf{\Gamma}_W \mathbf{S}(\boldsymbol{\omega})\|_F^2 = \left\| \mathbf{Y}_W - \sum_{m=1}^{M} \mathbf{\Gamma}_m \mathbf{s}(\omega_m) \right\|_F^2. \tag{5.6}$$

The RELAX idea is now simply to minimize (5.6) sequentially with respect to one signal at the time, while keeping the parameters of the other signals fixed. For the l-th signal we thus solve

$$\{\hat{\mathbf{\Gamma}}_l, \hat{\omega}_l\} = \arg\min_{\mathbf{\Gamma}_l, \omega_l} \|\mathbf{Y}_l - \mathbf{\Gamma}_l \mathbf{s}(\omega_l)\|_F^2, \tag{5.7}$$

where

$$\mathbf{Y}_l = \mathbf{Y}_W - \sum_{m \neq l} \hat{\mathbf{\Gamma}}_m \mathbf{s}(\hat{\omega}_m) \tag{5.8}$$

and $\hat{\mathbf{\Gamma}}_m$, $\hat{\omega}_m$, $m \neq l$, are the current estimates of the other targets' signal parameters. Solving (5.7) explicitly with respect to $\mathbf{\Gamma}_l$ results in an optimization of the single parameter ω_l only:

$$\hat{\omega}_l = \arg\max_{\omega_l} \|\mathbf{Y}_l \mathbf{s}^H(\omega_l)\|^2 \tag{5.9}$$

$$\hat{\mathbf{\Gamma}}_l = \mathbf{Y}_l \mathbf{s}^H(\hat{\omega}_l)/N\,. \tag{5.10}$$

Noting that the vector $\mathbf{Y}_l \mathbf{s}^H(\hat{\omega}_l)$ is composed of the DFT of the rows of $\mathbf{Y}_l$, the peak of $\|\mathbf{Y}_l \mathbf{s}^H(\omega_l)\|^2$ can efficiently be localized using FFT (with zero padding), followed by a local optimization. The whole procedure is initialized by assuming that the number of targets is $\tilde{M} = 1$ and solving for $\hat{\mathbf{\Gamma}}_1$ and $\hat{\omega}_1$. Next, the number of anticipated targets is increased to $\tilde{M} = 2$, and the RELAX search computes $\hat{\mathbf{\Gamma}}_1$, $\hat{\omega}_1$ and $\hat{\mathbf{\Gamma}}_2$, $\hat{\omega}_2$ iteratively until convergence. Next, $\tilde{M} = 3$ and so on, until the correct number of targets $\tilde{M} = M$ is reached. It is clear that such an approach fits very well with a sequential detection test for determining the number of targets, see e.g. [Eri02].

The full RELAX procedure is summarized as follows:

1. Initialize by $\tilde{M} = 1$ and $\mathbf{Y}_1 = \mathbf{Y}_W$.

2. Compute $\hat{\omega}_{\tilde{M}}$ and $\hat{\mathbf{\Gamma}}_{\tilde{M}}$ from (5.9)–(5.10).

3. For $\tilde{M}' = 2, 3, \ldots, \tilde{M}$, execute Step 4:

4. For $l = 1, 2, \ldots, \tilde{M}'$, compute

$$\begin{aligned} \mathbf{Y}_l &= \mathbf{Y}_W - \sum_{m=1, m\neq l}^{\tilde{M}'} \hat{\mathbf{\Gamma}}_m \mathbf{s}(\hat{\omega}_m) \\ \hat{\omega}_l &= \arg\max_{\omega_l} \|\mathbf{Y}_l \mathbf{s}^H(\omega_l)\|^2 \\ \hat{\mathbf{\Gamma}}_l &= \mathbf{Y}_l \mathbf{s}^H(\hat{\omega}_l)/N\,. \end{aligned}$$

   Repeat this step until convergence.

5. If $\tilde{M} < M$; let $\tilde{M} = \tilde{M} + 1$, compute $\mathbf{Y}_{\tilde{M}}$ according to (5.8) and goto 2.

The above scheme only involves local optimization over a single parameter. The convergence is usually substantially faster by replacing Steps 3-4 by a Newton-type search over the full parameter vector $[\omega_1, \ldots, \omega_{\tilde{M}}]^T$. However, the probability of finding the global optimum can sometimes be increased by doing at least one iteration of Steps 3-4 before switching to a local optimization.

Chapter 6

Numerical Examples and Simulations

In this chapter we present some numerical examples and results of Monte Carlo simulations. The purpose is twofold. On the one hand we want to corroborate the theoretical analysis presented in the preceding sections. On the other hand the theoretical analysis says nothing about how large N or K have to be in order to achieve the asymptotic regime. We will present some representative examples that will provide insight into this issue.

6.1 Cramér-Rao Bounds

First we will study how the 2-D and 1-D CRBs compare to the asymptotic bound as the number of samples increases. We will only consider the case of increasing number of pulses since the dual case of increasing number of antennas is obtained by replacing the roles of DOA and Doppler.

In the first example we study a scenario with two sources in spatio-temporally white Gaussian noise with variance $\sigma^2 = 1$. The source parameters are $\theta_1 = 20°, \theta_2 = 23°, \omega_1 = 0.40, \omega_2 = 0.55, b_1 = b_2 = 0.3 + j0.5$. The array employed is a 16 element ULA with $\lambda/2$ element separation. The null-to-null beamwidth of this array is $\Delta\theta \approx 14°$ and the Rayleigh resolution limit is thus 7°. Figure 6.1 shows the different CRBs on the DOA root mean square errors (RMSEs) as a function of the number of pulses N. The 1-D MLE cannot reach the 1-D CRB as $N \to \infty$ since the number of parameters grows with N. Therefore, we also plot the asymptotic variance of the MLE as given in [SN90a]. We observe that the 2-D CRB given by (3.11) coincides with the asymptotic or, equivalently, the single-source CRB given by (3.21) when N is sufficiently large. The number of required pulses to resolve the sources in Doppler according to the Rayleigh resolution limit is

$2\pi/(\omega_2 - \omega_1) \approx 42$. We observe from the plot that this is roughly the point where the 2-D CRB enters the asymptotic region and achieves the single-source bound. The 1-D CRB given by (4.2), however, does not attain the single-source bound no matter how large N is. The source DOA separation is much less than the spatial Rayleigh resolution limit and increasing N does not improve the spatial resolution. In this case the achievable performance when employing the 2-D model is much better than with a single 1-D model.

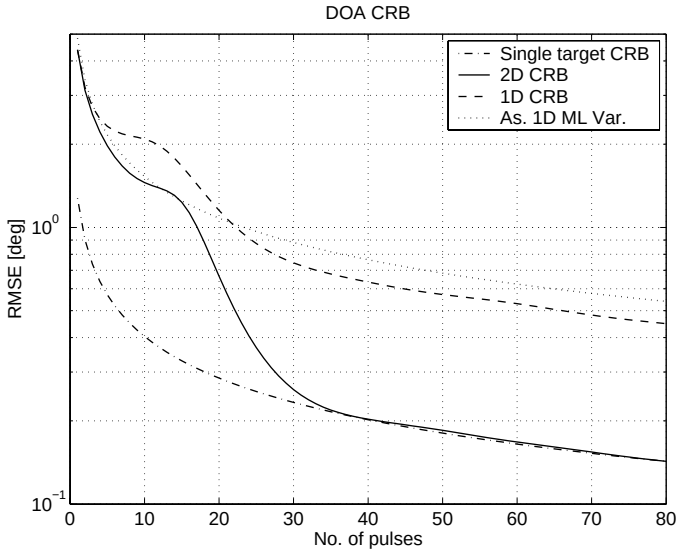

Figure 6.1: DOA CRBs as a function of the number of pulses, two sources and a 16 element ULA. The model parameters are $\sigma^2 = 1$, $\theta_1 = 20°, \theta_2 = 23°, \omega_1 = 0.40, \omega_2 = 0.55, b_1 = b_2 = 0.3 + j0.5$.

As we have previously seen, estimating with a two-step 1-D/1-D model can achieve the same asymptotic performance as with the full 2-D model, but only if the estimates are obtained in the right order. Since N is increasing in the present case, the Doppler parameters must be estimated first. The Doppler CRBs are plotted in Figure 6.2. In this case also the 1-D CRB reaches the single-source bound since the Doppler resolution is improved when increasing N. The corresponding MLE will reach the CRB as $N \to \infty$ since the number of parameters does not grow with N in the 1-D Doppler model.

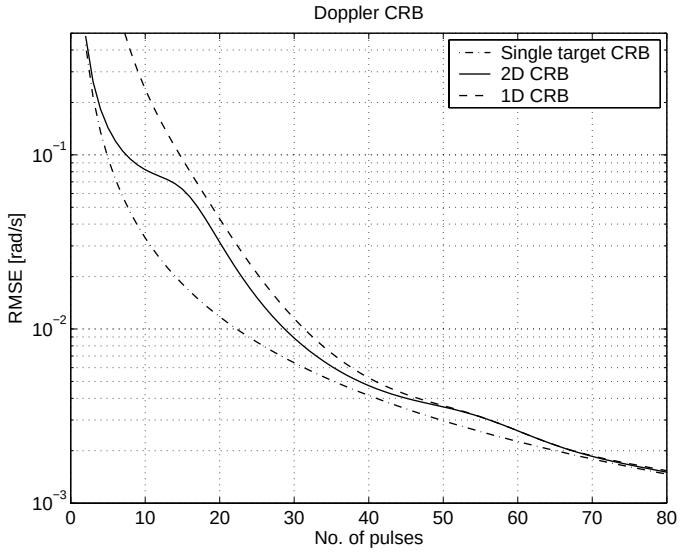

Figure 6.2: Doppler CRBs as a function of the number of pulses, two sources and a 16 element ULA. The model parameters are $\sigma^2 = 1$, $\theta_1 = 20°, \theta_2 = 23°, \omega_1 = 0.40, \omega_2 = 0.55, b_1 = b_2 = 0.3 + j0.5$.

6.2 Performance of the WLS Estimators

We now turn our attention to the actual performance of the 2-D and 1-D/1-D WLS estimators described previously. Since the finite sample performance is not available in analytical form we have to resort to Monte Carlo simulations.

We will study an example where we employ a ULA with 8 antenna elements spaced one-half wavelength apart. The null-to-null beamwidth of this array is 29° and the Rayleigh resolution limit is thus 14.5°. There are two sources present with parameters $\theta_1 = 18°, \theta_2 = 24°, \omega_1 = 0.20, \omega_2 = 0.10, b_1 = b_2 = 2$. The 2-D and 1-D/1-D WLS estimators were implemented and their RMSEs as a function of the number of pulses N were computed by means of Monte Carlo simulations. In order to avoid the effect of getting caught in a local maximum in search for the global maximum, we first initialized the search with the true source parameters. In practice, these are of course not known and some other initialization must be used. In order to compare with a more practical method we also implemented the RELAX procedure for the 1-D/1-D WLS estimator described in Section 5.2. The 2-D WLS estimator was then initialized with the estimates obtained from the 1-D/1-D WLS estimator. The local optimization was performed using a Gauss-Newton search. First we consider the case of spatio-temporally white noise and then study the case of spatially colored and temporally white noise. In the latter case,

the spatial color is produced by a directional interference (jammer).

White Noise

The results for spatio-temporally white Gaussian noise with variance $\sigma^2 = 1$ are depicted in Figures 6.3 and 6.4. Figure 6.3 shows the DOA estimation performance of the WLS estimators versus the number of pulses N. The pertinent CRBs are also shown in the plot. The 2-D WLS initialized with true parameter values provides excellent estimates; the RMSE is close the 2-D CRB for all N. The 1-D/1-D WLS estimator is equally good for large N but less accurate when N is small. We can observe a threshold below which the RELAX method failed to produce initial values within the region of attraction of the global optimum. For large N, however, the resulting estimation performance is the same as when initializing with the true parameter values. Indeed, all the estimators reach the single-source CRB when N is sufficiently large. It is interesting to note that this occurs when $N \approx 60$ which corresponds to the Doppler Rayleigh resolution limit being equal to the source Doppler separation. Figure 6.4 shows the results for the Doppler

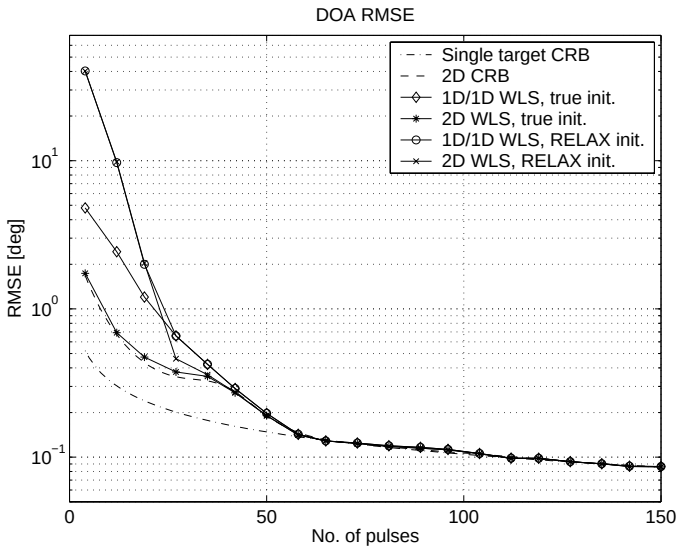

Figure 6.3: DOA estimation performance and CRBs versus the number of pulses, two sources, 8 element ULA, white noise. The source parameters are $\theta_1 = 18°, \theta_2 = 24°, \omega_1 = 0.20, \omega_2 = 0.10, b_1 = b_2 = 2$.

estimates. The relative performances between the estimators are the same as for the DOA estimates.

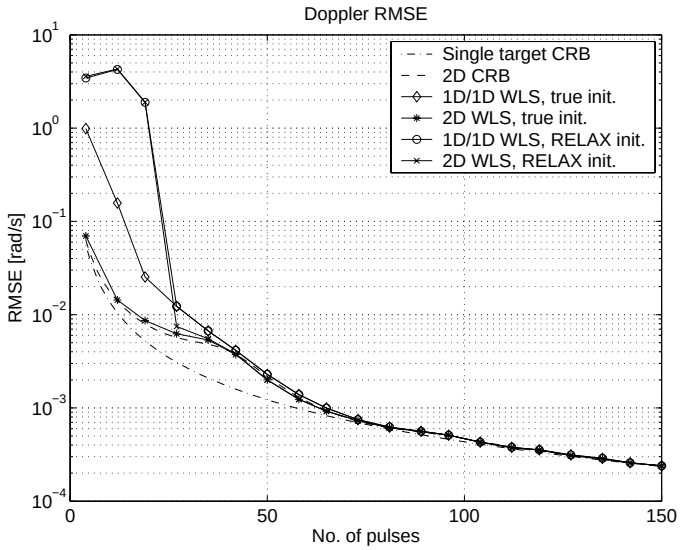

Figure 6.4: Doppler estimation performance and CRBs versus the number of pulses, two sources, 8 element ULA, white noise. The source parameters are $\theta_1 = 18°, \theta_2 = 24°, \omega_1 = 0.20, \omega_2 = 0.10, b_1 = b_2 = 2$.

Spatially Colored Noise

Next we consider the case of spatially colored and temporally white noise. We use the same scenario as in the white noise case with the exception of the presence of a directional interference with a DOA of $\theta_j = 35°$. The interference signal is zero mean, temporally white Gaussian noise with variance $\sigma_j^2 = 100$. The spatial noise covariance is thus $\mathbf{Q} = \sigma_j^2 \mathbf{a}(\theta_j)\mathbf{a}(\theta_j)^H + \mathbf{I}$.

The 2-D and 1-D/1-D WLS estimators were implemented using the 2-step approaches described in Subsections 3.5 and 4.3 respectively. Also in this case we considered the two cases of initialization with the true parameter values and the RELAX procedure. The DOA and Doppler estimation performance along with the CRBs are plotted versus the number of pulses, N, in Figures 6.5 and 6.6 respectively. The performance of the WLS estimators, relative to the CRBs, is somewhat worse for small N compared to the white noise case; whereas for large N the difference is negligible. This is expected since for small N the estimate of the noise covariance is poor. For large N the noise covariance is as good as known and the problem reduces to the white noise case by pre-whitening.

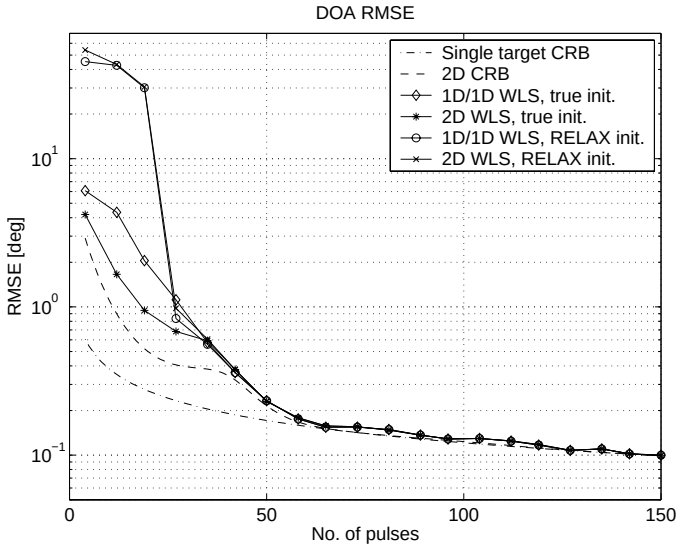

Figure 6.5: DOA estimation performance and CRBs versus the number of pulses, two sources, 8 element ULA. The source parameters are $\theta_1 = 18°, \theta_2 = 24°, \omega_1 = 0.20, \omega_2 = 0.10, b_1 = b_2 = 2$. White noise plus a jammer at $\theta_j = 35°$ with power $\sigma_j^2 = 100$.

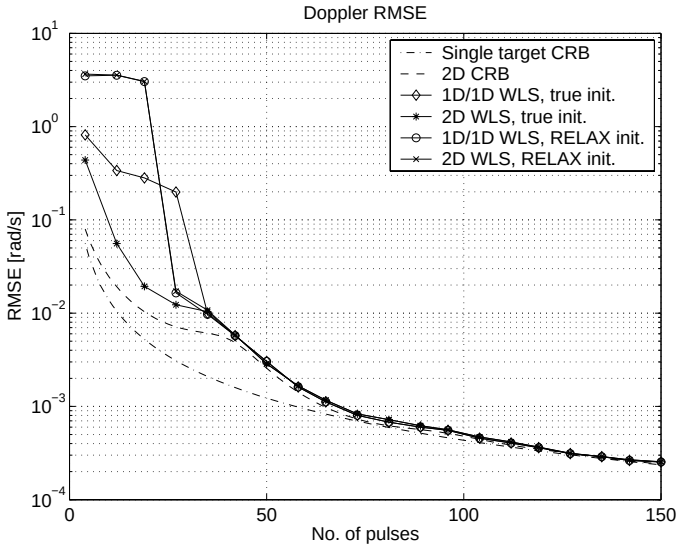

Figure 6.6: Doppler estimation performance and CRBs versus the number of pulses, two sources, 8 element ULA. The source parameters are $\theta_1 = 18°, \theta_2 = 24°, \omega_1 = 0.20, \omega_2 = 0.10, b_1 = b_2 = 2$. White noise plus a jammer at $\theta_j = 35°$ with power $\sigma_j^2 = 100$.

Chapter 7

Conclusions

In the first part of the thesis we have studied the problem of DOA and Doppler estimation with a pulsed Doppler radar system employing an array of antennas. This is an application of 2-D frequency estimation with DOA and Doppler corresponding to spatial and temporal frequencies, respectively. The focus has been on comparing the performance of joint and decoupled MLEs and the corresponding CRBs. We have shown that the joint DOA/Doppler problem decouples if the number of pulses or antenna elements is sufficiently large. Not only do the frequency dimensions decouple, but also the different sources. Thus, single-source performance is achievable in multi-source scenarios, if the amount of data is sufficiently large in at least one of the frequency dimensions. In order to achieve this, the estimation must performed in the right order. The frequency dimension in which the amount of data is large must be processed first. In this dimension, the sources are resolvable and, subsequently, the estimation in the other frequency dimension can be performed for one source at a time. The result is a tremendous reduction in computational complexity as compared to the full 2-D frequency estimation problem. Yet, under the stated assumptions the performance loss is insignificant. Numerical examples and results of Monte Carlo simulations indicated that the mentioned decouplings are valid when the sources are resolvable in one dimension according to the Rayleigh resolution limit.

Part II
Performance Analysis of DOA Estimation in the Threshold Region

Chapter 8

Introduction

The first part of this thesis deals with asymptotic analysis of DOA and Doppler estimation with sensor arrays. The tools used in this analysis are CRBs and Taylor expansions. With these tools we are able to show that the ML estimator attains the CRB if the amount of measured data is sufficiently large. In practical applications there are of course constraints on how many sensors one can afford and how long time there is available for collecting data.

The cost of a sensor array processing system depends to a great extent on the number of sensors, both in terms of hardware and computational requirements. Therefore, it is desirable to keep the number of sensors to a minimum while maintaining the system requirements. Likewise, the data collection time is also a limited resource since a real-world environment is often non-stationary. For instance, in radar applications the different targets are moving, the interference environment is changing and the background clutter is spatially inhomogeneous. Therefore, one is often limited to short measuring times in order to obtain a stationary data set and to be able to track the changing environment. Another parameter that may violate the asymptotic assumptions is a low SNR. A radar system often operates at low SNR — typically small targets are to be detected in a heavy clutter background and severe interference scenario.

A few words about asymptotics are in order. Due to averaging of the noise and, in many cases, coherent processing gain of the signals, a long measuring time results in a high integrated SNR. Therefore we can interchangeably speak of the asymptotic region as having high SNR or many data samples. In general, one may say that the asymptotic assumptions are usually valid whenever the estimation error is small. However, as we will see later when analyzing the deterministic and stochastic DOA estimators, this equivalence is

not always clear-cut. See [Asy98] for a further discussion about asymptotics.

Although asymptotic analysis is useful in many aspects, it does not provide a full picture of the achievable estimation performance of a sensor array system. It is also important to obtain an understanding of the inherent limitations in an estimation problem under conditions of low SNR and/or small number of data samples. Indeed, it is well known that nonlinear estimators typically exhibit a so called *threshold effect*, i.e. there is a point where the estimator more or less breaks down and yields very large errors.

The system design parameters that govern where the threshold regime sets in are often different from the ones that determine the performance in the asymptotic region. As will be shown in the sequel, the threshold in the DOA estimation problem is determined by the sidelobe level in the array beampattern, whereas the asymptotic accuracy depends only on the shape of the mainlobe. In fact, there is a trade-off between these two parameters. The more the sensors are spread out, the narrower is the mainlobe, but at the expense of high sidelobes. Therefore, designing such a system based only on asymptotic analysis would result in poor threshold performance.

In this part of the thesis we will develop a model of the threshold region performance of the ML DOA estimator. Actually, the basic principles behind the approach adopted herein could be used to model the threshold behavior of any estimator that is the minimizer of some multimodal criterion function. We will treat the ML estimation problems for both the deterministic and stochastic signal models. In order to make the presentation clearer we start with the simplest case first, which is that of a single source. Then we extend the analysis to multiple sources. The basic ideas carry over to the multi-source problem in a straightforward manner, but the technical details become considerably more intricate. Before we present the analysis, let us first give a little more background on the ubiquitous threshold effect.

8.1 The Threshold Effect

The threshold effect of nonlinear estimators and detectors is a well known phenomenon, see e.g. [Woo53, WJ65, Tre68, RB74]. In Figure 8.1 we try to illustrate the threshold behavior of the ML DOA estimator. The plot shows the MSE of the ML estimator and the CRB as a function of the SNR on a log-log scale. We can distinguish three different regions of operation. For high SNR the ML estimator attains the CRB. This is called the *asymptotic region*. When decreasing the SNR, at some point the MSE starts to rapidly deviate from the CRB. We call this point the *threshold* and the range of SNRs around it we refer to as the *threshold region*. For very low SNR the ML estimates

are practically uniformly distributed across the entire parameter space and, thus, provide no information. We call this the *no information region*. In this region, the MSE is equal to the a priori variance.

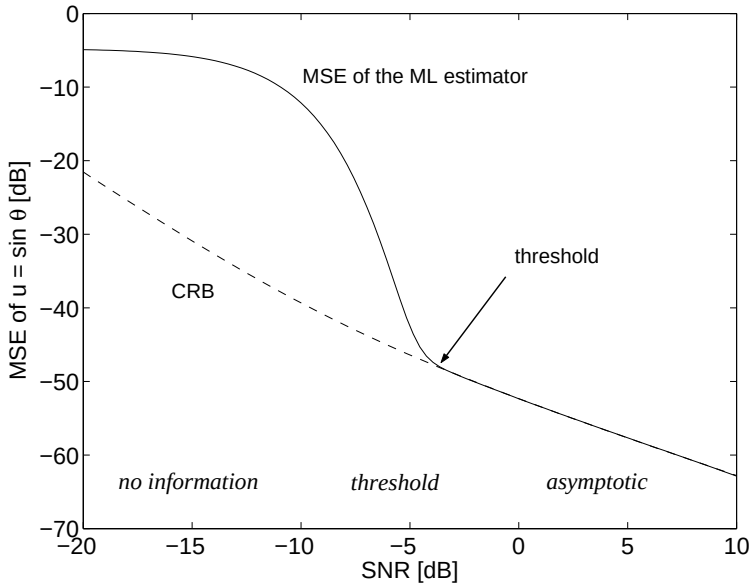

Figure 8.1: Different regions of operations of the ML DOA estimator.

Evidently, the CRB cannot predict the rapid deterioration of the ML performance below the threshold. This is due to the fact that the CRB is a local bound that depends essentially on the shape of the mainlobe in the ML criterion function, whereas the threshold effect is caused by large errors far away from the mainlobe. We call these large errors *outliers*. Figure 8.2 shows how an outlier may occur for the case of a single source in white Gaussian noise. The ML estimator is then equivalent to finding the maximum in the conventional beamformer spectrum. The left plot shows the array beampattern in the absence of noise. The beampattern has its global maximum at the true DOA which is zero. The right plot shows the beamformer spectrum when noise is present. In this case the global maximum is at a sidelobe peak; an outlier has occurred.

Clearly, the probability of outlier is related to the noise power and the sidelobe level in the array beampattern. Arrays having high sidelobes are more prone to deliver outliers than arrays with low sidelobes. Therefore, the threshold effect is more prominent for sparse arrays than for ULAs with a half-wavelength element separation. For a sparse array, the threshold can occur at a relatively high SNR. Thus, the threshold effect sets a limit to the degree of sparsity that can be tolerated. These issues are dealt with in more

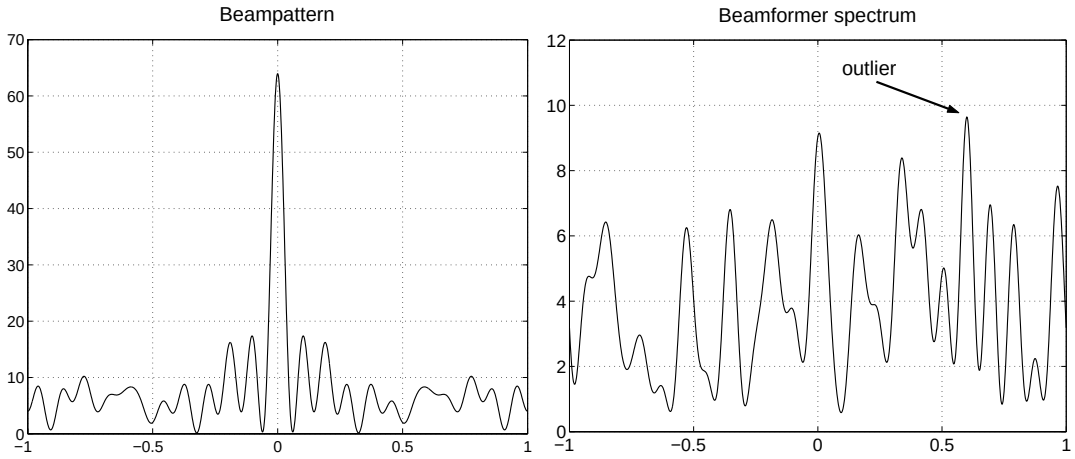

Figure 8.2: The left plot shows the array beampattern in the absence of noise. The right plot shows the beamformer spectrum when noise is present. This example is for an 8 element sparse array with element positions {0 1 2 11 15 18 21 23} $\lambda/2$, where λ is the wavelength.

detail in Chapter 11.

8.2 Performance Bounds

We have seen that the CRB cannot be used to predict the threshold effect of the ML estimator. As alluded to previously, the reason for this is that it is a local bound that only considers small deviations from the true parameter values. Several other lower bounds have been developed over the years that take the effect of large errors into account. With these bounds it is possible predict the threshold effect. Examples of such bounds are the Barankin [Bar49], Ziv-Zakai [ZZ69], and Weiss-Weinstein [WW85] bounds.

The Cramér-Rao and Barankin bounds belong to the "classical" bounds, which treat the parameters as unknown deterministic quantities. They provide lower bounds on the MSE of the parameter estimates for each possible value of the parameter. In order not to obtain the trivial bound of zero, some restriction must be imposed on the estimator. The restriction that these bounds impose is that the estimator must be unbiased.

Another class of bounds are Bayesian bounds, which assume that the parameters are random variables with known prior distributions. They provide bounds on the global MSE averaged over the prior distributions. There are no restrictions on the estimators which they apply to. In contrast to classical estimation theory, Bayesian theory provides an explicit solution for the min-

imum mean square error (MMSE) estimator, which is the conditional mean estimator [Kay93]. There is also a version of the CRB that applies to random parameters, called the Bayesian CRB, see [Tre68]. For a thorough treatment of Bayesian bounds in parameter estimation, see [Bel95].

A crucial issue with regard to using lower bounds for predicting the threshold behavior is if they are *tight*, i.e. if there actually exists an estimator that can attain the bound. The MSE of the Bayesian MMSE estimator is of course a tight lower bound under the Bayesian framework, and the original formulation of the Barankin bound is also the greatest lower bound under the appropriate assumptions. Unfortunately, these bounds are impractical to compute for all but the simplest of problems. Therefore, other lower bounds have been developed that are more practical to compute, such as the Weiss-Weinstein and Ziv-Zakai bounds, and a simpler form of the Barankin bound in [MH71].

Several authors have applied these bounds to the DOA estimation problem, see [Ver87, DeL93, NV94, Ngu94, Bel95, BET96, ZS98, AE00, Ath01b, BS01]. The Weiss-Weinstein and Ziv-Zakai bounds appear to give a reasonable prediction of the performance of the ML estimator in the threshold region for a single source. The simplified versions of the Barankin bound appears to be weaker. However, care has to be taken when interpreting Bayesian bounds, since they can be strongly influenced by the parameter values which produce the largest errors. This makes it difficult to apply these bounds to the multi-source DOA estimation problem. If, for example, two sources are present, the estimation error will be very large when the DOA separation is small whereas the error will be small if the separation is large. Since the Bayesian MSE is averaged over all parameter values, it will be dominated by the MSE for small source separations.

A way to overcome this problem is to condition on a particular DOA separation and evaluate the bound for this separation. However, this introduces information that is not normally available a priori which would lead to a weak bound. In [NV94, Ngu94] the Weiss-Weinstein bound was implemented for a two-source DOA estimation problem. Simulations showed that this bound was far too optimistic in predicting the MSE of the ML estimator in the threshold region. Another way could be to adjust the prior distribution in order to compensate for the strong dependence of the error on the parameter values. We will not pursue these issues any further in this thesis. The difficulty in applying these bounds to accurately estimate the threshold of the ML estimator for the two-source problem was also noted in [Tre02, Section 8.12.2].

We may identify two tracks to study the fundamental limitations in an estimation problem. On the one hand we can study different performance

bounds that are more or less tight. On the other hand we can study the performance of the best estimator that we know of. Usually, this estimator is the ML estimator. In this part of the thesis the main focus is on the performance of the ML estimator. The reasons for this are summarized in the following list

- Tight (and computable) lower bounds for the multi-source DOA estimation problem are not available in the literature.
- For the single source problem, there is still a gap between the best lower bound that we know of and the performance of the ML estimator.
- To the author's knowledge, no estimator with better threshold performance than the ML estimator is known. Therefore, the MSE of the ML estimator might serve as a "practical" lower bound. In Chapter 11 we will show an example in which the performance of the ML estimator is very close to that of the MMSE estimator.
- There seems to be no general analysis available in the literature that provides an accurate characterization of the threshold region performance of the ML DOA estimator.
- Developing models for the MSE of the ML estimator may provide new insights that are not obvious to interpret in lower bounds.
- Lower bounds like the Weiss-Weinstein and Barankin bounds are far from trivial to implement. For example, they require maximization over a set of "test points", and if these are not chosen judiciously one may encounter numerical problems.
- Bayesian bounds, which appear to be the tightest ones, require that prior distributions are assigned to all unknown parameters. It is not obvious how to assign such prior distributions.

Although the main focus of this part of the thesis is on modeling the MSE of the ML estimator, we will also, on several occasions, include discussions on bounds. It is interesting to compare the performance predictions the bounds provide with the MSE approximations that we will develop in the sequel.

8.3 Outline and Contributions

Part II of the thesis is organized as follows. First, we will develop a model for the performance of the ML DOA estimator for the single source case.

We then generalize this analysis to account for multiple sources. The thesis ends with two self-contained sections with application examples that provide further insights into the problem of DOA estimation in the threshold region. These examples deal with the single source problem.

In the first example, we study an interferometric-like DOA estimation system consisting of two widely separated subarrays. A discussion on Bayes-optimal DOA estimation is included and we argue that the ML estimator can be considered as "nearly" optimal for this problem also when taking the threshold effect into account.

In the second example, we use the Weiss-Weinstein lower bound to optimize the element positions in sparse arrays. Even if this bound is not perfectly tight, one might conjecture that the relevant system design parameters have similar impact on the actual performance as they have on the bound.

The main contribution of the second part of this thesis is that it provides an extensive study of the threshold properties of the ML DOA estimator. The asymptotic properties of this estimator are fairly well known. It has been shown that the stochastic ML (SML) estimator achieves the corresponding CRB as the number of snapshots tends to infinity in [SN90b]. In [Vib93] it was shown that the SML estimator when applied to deterministic signals reaches the deterministic CRB as the SNR tends to infinity. However, a high SNR analysis of the SML estimator when applied to stochastic signals seems to be lacking in the literature. As for the deterministic ML (DML) method, it was shown to be asymptotically efficient as the SNR [Vib93] or number of sensors [Vib95] tends to infinity. Furthermore, it has been shown that the DML estimator does not achieve the CRB as the number of snapshots tends to infinity, unless the number of sensors also tends to infinity [SN89]. The threshold region properties of the ML DOA estimators appears to be a less explored area. To quote [Asy98], the finite-sample (or SNR, etc.) behavior of most detection or estimation methods is *terra incognita*.

The basic idea behind the threshold region analysis presented herein dates back to [Tre68]. This concept was applied successfully in [RB74] for predicting the threshold behavior of ML frequency estimation of a single tone. Because of the duality between frequency and DOA estimation one can also apply the analysis in [RB74] to this problem. However, there are a couple of simplistic assumptions in [RB74] that are not appropriate for DOA estimation with sensor arrays, especially when nonuniform arrays are employed. Furthermore, their analysis is not applicable to the multi-source problem. In [Ver87, VH91], the approach of [RB74] was implemented for the single source DOA estimation problem, and was found to deliver overly optimistic performance predictions. An improvement was suggested by an empirical

modification of the method in [RB74]. For the particular problem setup in [Ver87, VH91] the results were accurate.

In this thesis, we provide a more theoretically well-founded approach than the one in [Ver87, VH91]. It is a refined version of the approach in [RB74] that is also applicable to nonuniform arrays and the multi-source problem. We published the results on the ML threshold analysis in [Ath02a, Ath02b, Ath02c]. Independently, and about the same time, a very similar analysis for a matched-field problem in sonar signal processing was published in [Xu01, Xu02]. An important difference is that [Xu01, Xu02] treats the case of one signal that has multiple parameters, whereas in [Ath02b] we treat the case of multiple sources and one parameter per source. We also treat the case of deterministic signals, which is not covered in [Xu01, Xu02]. Very recently, [Ric03] applied the same analysis to handle colored noise and secondary data for the case of a single deterministic signal. The results presented in Section 10.4 on the threshold analysis of the SML estimator for the multi-source problem have not been published elsewhere.

The main contribution of the application example in Section 11.1 is that it provides some further insights into DOA estimation using two separated subarrays. Several authors have treated various, more narrow, aspects of this problem. In [ZS98] the CRB was compared with the Weiss-Weinstein bound, in [LvdVD99, WZ98, ZW00] estimation algorithms specifically designed to resolve ambiguities were presented and in [Bom00] the ML estimator and the Barankin bound were computed for this array structure. This section is an extended version of [AE00], and a continuation of this work has been presented in [ES02].

Finally, in the last section, we propose a method for optimizing the element positions in sparse linear arrays. Different approaches to optimizing the element positions with respect to DOA estimation accuracy have been taken in the literature. In [AGGS96, GB97], the element positions of non-uniform linear arrays were optimized by minimization of the CRB. A problem with this approach is that high sidelobes have no effect on the CRB.

Various approaches have been proposed to account for the effect of high sidelobes. In [VE99] the mainlobe area was minimized subject to a peak sidelobe constraint and in [AS00] competitive criteria involving maximum aperture and identifiability were considered. Although these approaches are intuitively appealing, there is no explicit connection between these ambiguity/aperture trade-offs and the resulting MSE. Herein, we propose a novel criterion for optimizing the element positions of sparse linear arrays. The criterion used is the SNR threshold of the Weiss-Weinstein lower bound. In this way we take near ambiguities into account while having an immediate link to the attainable MSE. An interesting result is that the array obtained

from the optimization and the so called minimum redundancy array [Mof68] have nearly identical Weiss-Weinstein lower bounds for the single source case.

Chapter 9

ML DOA Estimation of a Single Source

In this chapter we will present an analysis of the threshold region performance of the ML DOA estimator for the single source case. We will consider this case in somewhat more detail than the multi-source problem. The reason for this is twofold:

1. It is easier to provide pedagogical clarity in the single source case since the multi-source problem is technically more awkward.

2. Validating the theoretical analysis with Monte Carlo simulations is very time-consuming for the multi-source problem. The single source problem permits us to study more aspects of the inherent limitations in the DOA estimation problem since time constraints are not that restrictive.

9.1 Data Model And Problem Formulation

Consider an array with K sensors receiving a single narrowband signal from a far-field source. For simplicity, assume that the array and the source are coplanar so that the DOA can be characterized by the azimuthal angle only. The complex baseband array output is modeled by the $K \times 1$ vector

$$\mathbf{x}(t) = \mathbf{a}(\theta_0)s(t) + \mathbf{n}(t), \qquad t = 1, \ldots, N \tag{9.1}$$

where $\mathbf{a}(\theta_0)$ is the $K \times 1$ array steering vector that models the array response to a unit waveform from the DOA θ_0, measured relative to the array boresight. The sensors are assumed to be omnidirectional with unity gain. The

array steering vector then takes the form

$$\mathbf{a}(\theta) = \left[\begin{array}{ccc} e^{-j\pi d_1 \sin\theta} & \ldots & e^{-j\pi d_K \sin\theta} \end{array} \right]^T, \tag{9.2}$$

where $d_k, k = 1, \ldots, K$ are the sensor positions in units of half wavelengths. Furthermore, N denotes the number of snapshots, $s(t)$ is the complex amplitude at baseband of the impinging wavefront and $\mathbf{n}(t)$ is an additive noise term. The noise $\mathbf{n}(t)$ is assumed to be a stationary, zero mean, complex Gaussian process with second order moments

$$\begin{aligned} \mathrm{E}[\mathbf{n}(t)\mathbf{n}^H(\tau)] &= \sigma_n^2 \mathbf{I}\, \delta_{t,\tau} & (9.3) \\ \mathrm{E}[\mathbf{n}(t)\mathbf{n}^T(\tau)] &= \mathbf{0} \quad \forall\, t, \tau, & (9.4) \end{aligned}$$

where $\delta_{t,\tau}$ is the Kronecker delta. Thus, the noise is assumed to be spatio-temporally white with variance σ_n^2.

Two different models for the signal, $s(t)$, will be considered; the stochastic and deterministic models. In the stochastic model, the signal is assumed to be a stationary, zero mean, complex Gaussian process with second order moments

$$\begin{aligned} \mathrm{E}\left[s(t)s^*(\tau)\right] &= \sigma_s^2 \delta_{t,\tau} & (9.5) \\ \mathrm{E}\left[s(t)s(\tau)\right] &= 0 \quad \forall\, t, \tau. & (9.6) \end{aligned}$$

The signal is thus assumed temporally white with variance σ_s^2. Furthermore, it is assumed that the signal and the noises are statistically independent. In the deterministic model, the signal samples $\{s(t)\}_{t=1}^N$ are regarded as unknown, deterministic parameters that have to be estimated. The signal is considered nonrandom, i.e. the sequence $\{s(t)\}_{t=1}^N$ is frozen in all realizations of $\{\mathbf{x}(t)\}_{t=1}^N$, only $\{\mathbf{n}(t)\}_{t=1}^N$ varies. Under the stochastic model assumption, the signal is random, *i.e.* also the sequence $\{s(t)\}_{t=1}^N$ varies from realization to realization. One may also think of the deterministic model as having conditioned on a particular realization of a random $s(t)$. Therefore, the models are also sometimes referred to as unconditional and conditional models, see [SN90b].

The problem considered herein is; given the noisy measurements $\mathbf{x}(t), t = 1, \ldots, N$, to estimate the DOA of the far-field source. For mathematical convenience, the estimation of $u_0 \triangleq \sin\theta_0$ will be considered. More specifically, our aim is to develop analysis tools that are able to accurately characterize the performance of the ML estimator in the threshold region, where traditional tools such as the CRB are too optimistic. When designing a DOA estimation system, it is usually desired to operate above threshold. It is therefore important to be able to determine whether a specific design will

make the system operate above threshold or not. This is of particular interest in sparse array design where the threshold effect is more prominent, since such arrays will have high sidelobes due to the spatial undersampling.

It is easy to show that ML estimation of u_0, under the stochastic *or* deterministic model assumption, amounts to finding the maximum of the beamformer spectrum[1]

$$V(u) = \frac{1}{N}\sum_{t=1}^{N} |\mathbf{a}^H(u)\mathbf{x}(t)|^2 = \mathbf{a}^H(u)\hat{\mathbf{R}}\mathbf{a}(u), \tag{9.7}$$

where $\hat{\mathbf{R}} = \frac{1}{N}\sum_{t=1}^{N}\mathbf{x}(t)\mathbf{x}^H(t)$ is the sample covariance matrix and $u \triangleq \sin\theta$. The ML estimator is thus the same under the stochastic and deterministic models for the case of a single source in white Gaussian noise. However, the performance analysis will be different since the measured data in $\hat{\mathbf{R}}$ have different statistical properties under the two models.

A property of the array that is intrinsic to the performance of any DOA estimation algorithm is the *beampattern*, defined by

$$g(u, u_0) = \mathbf{a}^H(u)\mathbf{a}(u_0) = \sum_{k=1}^{K} e^{j\pi d_k(u-u_0)}. \tag{9.8}$$

The beampattern thus represents the complex gain as a function of u of a beamformer steered toward the direction u_0. Usually, we are mostly interested in the modulus of the beampattern. Therefore, it should be understood that whenever we speak of the beampattern in the sequel, we mean $|g(u, u_0)|$. The beampattern is characterized by a mainlobe and a number of sidelobes. The beamformer spectrum can be seen as a noise-corrupted version of the beampattern. When the SNR is high, the ML estimate will be close to the beampattern's mainlobe peak. At low SNR, however, the noise may cause a sidelobe to become of greater magnitude than the mainlobe. In this case, the ML estimator will pick the sidelobe peak as the estimate, yielding a large estimation error. Such estimates are termed outliers and they are the cause of the threshold effect.

9.2 Approximation of the MSE

In this section, an approximation to the MSE of the ML estimator that takes the threshold effect into account is derived. The basic idea behind the

[1]For the deterministic model it is trivial (just replace $\mathbf{A}$ with $\mathbf{a}$ in (1.53)). For the stochastic model, this requires that $|\mathbf{a}(u)|$ is independent of u, see [Vib95]. Evidently, this holds for the steering vector in (9.2). In Section 11.1.4 we show it for the case of known SNR. However, it turns out that the form of the estimator is independent of the SNR.

approximation is to divide the MSE into two parts. The first part is due to small errors close to the true value. These local errors are well described by standard analysis tools, such as the CRB and Taylor expansions. The second part is due to outliers from the sidelobe region. Since the CRB and Taylor expansions are local in nature, large errors from outliers will not be captured by such an analysis.

By the total probability theorem, the MSE can be written as a sum of two parts according to

$$\begin{aligned} \mathrm{E}\left[(\hat{u}-u_0)^2\right] &= \Pr[\text{ no outlier }]\,\mathrm{E}\left[(\hat{u}-u_0)^2|\text{ no outlier }\right] \\ &+ \Pr[\text{ outlier }]\,\mathrm{E}\left[(\hat{u}-u_0)^2|\text{ outlier }\right], \end{aligned} \tag{9.9}$$

where "outlier" denotes the event that the global maximum in the beamformer spectrum is outside the mainlobe area (e.g. the null-to-null beamwidth). The idea to divide the MSE in this manner is not new, see e.g. [Tre68, RB74]. In [Tre68, RB74] it was assumed that the distribution of the estimates is uniform when an outlier occurs, since the signal has little influence. Although this is a reasonable assumption in the no information region, it is a crude approximation in the threshold region. In Figure 9.1 we plot the beampattern of a four element minimum redundancy array (MRA) together with a histogram of $2 \cdot 10^5$ ML estimates for an SNR value in the threshold region. Clearly, the histogram has the same shape as the beampattern. It should therefore be possible to obtain a more accurate approximation of the MSE in the threshold region if knowledge of the beampattern was incorporated.

In order to obtain a more refined, and yet simple, approximation of the MSE and probability of outlier we approximate the continuous beampattern with a discretized version according to

$$\tilde{g}(u,u_0) = \sum_{n=0}^{N_p} g(u_n,u_0)\delta(u-u_n). \tag{9.10}$$

A reasonable approximation is to sample the continuous beampattern at the peaks. We then let u_0 be the mainlobe peak and $u_n, n = 1,\ldots,N_p$, be the sidelobe peaks. Invoking the total probability theorem again, we obtain for the second term in (9.9)

$$\begin{aligned} &\Pr[\text{ outlier }]\,\mathrm{E}\left[(\hat{u}-u_0)^2|\text{ outlier }\right] = \\ &\sum_{n=1}^{N} \Pr[\text{ sidelobe } n]\,\mathrm{E}\left[(\hat{u}-u_0)^2|\text{ sidelobe } n\right], \end{aligned} \tag{9.11}$$

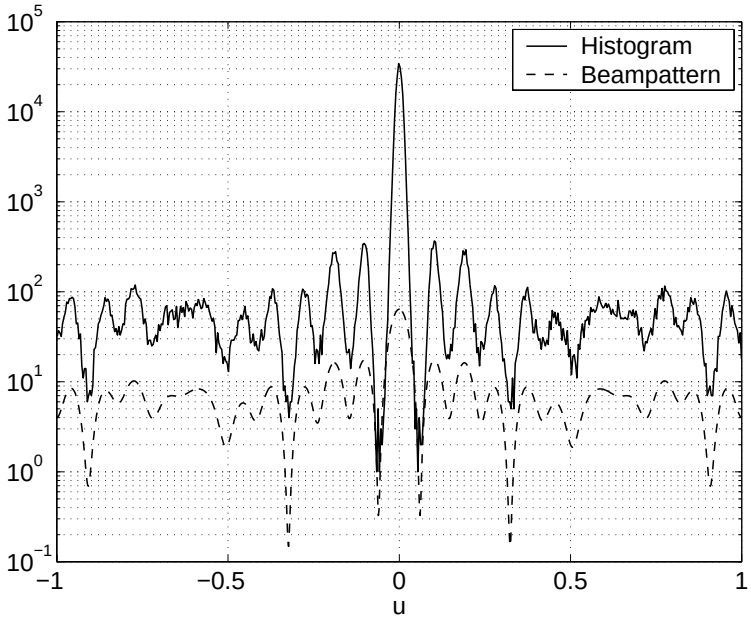

Figure 9.1: Beampattern for a 4 element MRA and histogram of DOA estimates. Deterministic model, 1 snapshot, SNR = 1 dB.

where the notation "sidelobe n" means the event that the n-th sidelobe peak in the discretized beamformer spectrum is of greater magnitude than all other peaks. The conditional MSE when an outlier comes from the n-th sidelobe peak is simply given by

$$\mathrm{E}\left[(\hat{u}-u_0)^2|\text{ sidelobe } n\right] \;=\; (u_n-u_0)^2. \tag{9.12}$$

This, in effect, means that we have ignored the local variations around the sidelobe peaks.

In order to obtain a good approximation to the MSE also at high SNR, the local errors around the true direction need to be considered. To achieve this, the MSE conditioned on the event that no outliers have occurred will be approximated by the CRB. The approximation to the total MSE thus becomes

$$\mathrm{E}\left[(\hat{u}-u_0)^2\right] \;=\; \Pr[\text{ no outlier }]\;\mathrm{CRB} + \sum_{n=1}^{N_p} \Pr[\text{ sidelobe } n]\;(u_n-u_0)^2. \tag{9.13}$$

This approximation is actually quite general since nothing specific to the ML estimator has been used in its derivation. It could be applied to any

estimation algorithm that finds its estimates by searching for the maximum (or minimum) of some cost function. If the estimator is not asymptotically efficient, the CRB in (9.13) should be replaced with the asymptotic variance of the estimator (if known). The problem specific quantities are the outlier probabilities, which will depend on the statistics of the cost function, and the CRB which will depend on the data model.

The CRB for the deterministic model has been derived in [SN89]. For a single source (and the assumptions in Section 2) this can be written as

$$\mathrm{CRB} = \frac{1}{2\pi^2 SU}, \tag{9.14}$$

where

$$S \triangleq \frac{K}{\sigma_n^2} \sum_{t=1}^{N} |s(t)|^2 \tag{9.15}$$

$$U \triangleq \frac{1}{K} \sum_{m=1}^{K} \left(d_m - \frac{1}{K} \sum_{n=1}^{K} d_n \right)^2 . \tag{9.16}$$

Here, S may be interpreted as the total SNR integrated over the spatial and temporal samples and U as the variance of the element positions. The stochastic CRB has been derived in [SN90b], which for a single source can be expressed as

$$\mathrm{CRB} = \frac{K \cdot \mathrm{SNR} + 1}{2\pi^2 N K^2 \mathrm{SNR}^2 U}, \tag{9.17}$$

where

$$\mathrm{SNR} \triangleq \sigma_s^2 / \sigma_n^2. \tag{9.18}$$

Note that the deterministic and stochastic CRB coincide for high SNR if one lets $\sigma_s^2 = \frac{1}{N} \sum_{t=1}^{N} |s(t)|^2$.

9.3 Probability of Outlier

This section presents derivations of the outlier probabilities needed for obtaining the MSE approximation in (9.13). These probabilities also have an interest in their own, since the MSE may not be the only performance measure to consider. In fact, in some applications the probability of outlier may be a more relevant measure than the MSE. One might consider an outlier as an unacceptable event and that the actual size of the error, in this event,

is immaterial. In such a case, a measure of local accuracy (e.g. the CRB) together with the probability of outlier may provide more adequate information. Still, it is interesting to obtain an accurate MSE approximation since, by predicting the threshold, we can tell under which conditions we can rely on asymptotic results. Herein, we develop models for all these quantities and it is the user's choice which one to pick for a particular application.

Now, the probability of outlier is the probability of the union of events that one of the sidelobe peaks is higher than the mainlobe peak, i.e.

$$P_o = \Pr\left[\bigcup_{n=1}^{N_p} \{V(u_n) > V(u_0)\}\right], \tag{9.19}$$

where P_o denotes the probability of outlier. Since the events $\{V(u_n) > V(u_0)\}$ are not disjoint, exact computation of this probability is a difficult, if not impossible, task. However, recall the following rule from elementary probability theory [GS92]

$$\begin{aligned}\Pr\left[\bigcup_{n=1}^{N_p} A_n\right] &= \sum_n \Pr[A_n] - \sum_{n<m} \Pr[A_n \cap A_m] + \sum_{n<m<k} \Pr[A_n \cap A_m \cap A_k] \\ &+ (-1)^{N_p+1} \Pr[A_1 \cap A_2 \cap \cdots \cap A_{N_p}],\end{aligned} \tag{9.20}$$

where $A_n, n = 1, \ldots, N_p$, are events and $\sum_{n<m}$ sums over all unordered pairs (n, m) for $n < m$. Actually, the probability on the left hand side is upper bounded by the first sum on the right hand side. This inequality is in the literature referred to as the *union bound* or *Boole's inequality*.

If we let A_n be the event $\{V(u_n) > V(u_0)\}$, we see that the other sums in (9.20) contain probabilities of events that two or more sidelobe peaks simultaneously are higher than the mainlobe peak. At high SNR it is very unlikely that a sidelobe peak is higher than the mainlobe peak. The event that two or more sidelobe peaks simultaneously are higher than the mainlobe peak is then even more unlikely. In this case, the first sum in (9.20) would be a good approximation to the probability on the left hand side. In fact, as we will demonstrate later, this approximation is good also in the threshold region. Not until the SNR is so low that we are in the no information region, does the approximation degrade. Since the estimator is more or less useless anyway in this region, we have no interest in modeling the performance under such conditions. Therefore, we will not try find a more accurate approximation. Hence, we approximate the probability of outlier in (9.19) by

$$P_o \approx \sum_{n=1}^{N_p} P_n, \tag{9.21}$$

where

$$P_n \triangleq \Pr\left[V(u_n) > V(u_0)\right]. \tag{9.22}$$

We call the approximation in (9.21) the *union bound approximation* of the probability of outlier.

Let us digress briefly and note an interesting connection to detection theory. We can interpret our estimation problem as an N_p-ary detection problem commonly encountered in e.g. communication theory. The peaks in the beampattern then correspond to receiver outputs matched to different signal waveforms. The mainlobe peak corresponds to the output of the receiver matched to the transmitted signal. Furthermore, the probability of outlier in the estimation problem corresponds to the probability of error in the detection problem. The individual probabilities, P_n, are in communication theory often referred to as the *pairwise error probabilities*. We will also use this term in the sequel when referring to the probabilities in our estimation problem.

The next probability that needs to be computed is Pr[sidelobe n], i.e. the probability that the global maximum in the discretized beamformer spectrum is at the n-th sidelobe peak in the beampattern. This probability is also difficult to compute exactly. Again, a union bound will be used to approximate this probability. The sought probability can be expressed as

$$\Pr[\text{ sidelobe } n] = 1 - \Pr\left[\bigcup_{\substack{k=0\\k\neq n}}^{N_p} \{V(u_k) > V(u_n)\}\right]. \tag{9.23}$$

Now, the probability on the right hand side is approximated by its union bound according to

$$\begin{aligned}\Pr\left[\bigcup_{\substack{k=0\\k\neq n}}^{N_p} \{V(u_k) > V(u_n)\}\right] &\approx \sum_{\substack{k=0\\k\neq n}}^{N_p} \Pr\left[V(u_k) > V(u_n)\right] \\ &\approx \Pr\left[V(u_0) > V(u_n)\right],\end{aligned} \tag{9.24}$$

where only the dominant term in the sum has been retained. This term is the probability that the mainlobe is higher than sidelobe n, which is much higher than the probabilities that one of the other sidelobes is higher than sidelobe n. Using these approximations, we obtain

$$\Pr[\text{ sidelobe } n] \approx 1 - \Pr\left[V(u_0) > V(u_n)\right] = P_n. \tag{9.25}$$

Thus, the problem has been reduced to computing the pairwise error probabilities, $P_n, n = 1, \ldots, N_p$. Returning to the MSE approximation in (9.13), this can now be written as

$$\mathrm{E}\left[(\hat{u} - u_0)^2\right] \approx \left(1 - \sum_{n=1}^{N_p} P_n\right) \mathrm{CRB} + \sum_{n=1}^{N_p} P_n \, (u_n - u_0)^2. \tag{9.26}$$

This is the final expression for MSE approximation that will be used in the rest of this chapter.

In order to compute P_n, let y_0 and y_n denote the beamformer outputs when steered to the directions u_0 and u_n respectively, *i.e.*

$$\begin{aligned} y_0(t) &= \mathbf{a}^H(u_0)\mathbf{x}(t) = Ks(t) + \mathbf{a}_0^H\mathbf{n}(t) && (9.27)\\ y_n(t) &= \mathbf{a}^H(u_n)\mathbf{x}(t) = \mathbf{a}_n^H\mathbf{a}_0 s(t) + \mathbf{a}_n^H\mathbf{n}(t), && (9.28) \end{aligned}$$

where, for brevity, $\mathbf{a}_0 \triangleq \mathbf{a}(u_0)$ and $\mathbf{a}_n \triangleq \mathbf{a}(u_n)$. Using (9.7), (9.22), (9.27) and (9.28) we obtain

$$P_n = \Pr\left[\sum_{t=1}^{N} \left(|y_0(t)|^2 - |y_n(t)|^2\right) < 0\right]. \tag{9.29}$$

Since y_0 and y_n are linear transformations of the same complex, Gaussian random vector $\mathbf{x}$, they are jointly Gaussian random variables. The transformation is not orthogonal (except when u_n is at a null in the beampattern), which means that y_0 and y_n are correlated. In order to compute this probability, we will use a result from Appendix B in [Pro95]. Therein, the probability that a Hermitian quadratic form in complex Gaussian random variables is less than zero is computed. The sum in (9.29) is a special case of such a quadratic form. The moments of y_0 and y_n will be different under the deterministic and stochastic model assumptions, giving different probabilities. The next two subsections will present the resulting probabilities for the two models.

9.3.1 Deterministic Signal Model

Under the deterministic signal model, the means of y_0 and y_n are

$$\begin{aligned} \mathrm{E}[y_0] &= Ks(t) && (9.30)\\ \mathrm{E}[y_n] &= \mathbf{a}_n^H\mathbf{a}_0 s(t), && (9.31) \end{aligned}$$

while the second order moments are given by

$$\begin{aligned} \mathrm{Var}[y_0] &= \mathrm{Var}[y_n] = K\sigma_n^2 && (9.32)\\ \mathrm{Cov}[y_0, y_n] &= \mathbf{a}_0^H\mathbf{a}_n\sigma_n^2. && (9.33) \end{aligned}$$

Applying the results from [Pro95], we obtain the following expression for P_n after some straightforward algebra

$$\begin{aligned} P_n &= Q\left(\sqrt{\frac{S}{2}\left(1-\sqrt{1-r_n^2}\right)}, \sqrt{\frac{S}{2}\left(1+\sqrt{1-r_n^2}\right)}\right) \\ &- e^{-S/2}\left[I_0\left(\frac{r_n S}{2}\right) - \frac{1}{2^{2N-1}} I_0\left(\frac{r_n S}{2}\right) \sum_{m=0}^{N-1} \binom{2N-1}{m}\right. \\ &- \frac{1}{2^{2N-1}} \sum_{m=1}^{N-1} I_m\left(\frac{r_n S}{2}\right) \left(\left(\frac{1+\sqrt{1-r_n^2}}{r_n}\right)^m - \left(\frac{1-\sqrt{1-r_n^2}}{r_n}\right)^m\right) \\ &\times \left.\sum_{k=0}^{N-1-m} \binom{2N-1}{k}\right], \end{aligned} \tag{9.34}$$

where

$$Q(\alpha,\beta) = \int_\beta^\infty t e^{-(t^2+\alpha^2)/2} I_0(\alpha t)\, dt \tag{9.35}$$

is Marcum's Q function and

$$I_m(x) = \frac{(x/2)^m}{\sqrt{\pi}\Gamma(m+\frac{1}{2})} \int_0^\pi e^{x\cos\theta} \sin^{2m}\theta\, d\theta = \sum_{k=0}^\infty \frac{(x/2)^{2k+m}}{k!\Gamma(m+k+1)} \tag{9.36}$$

is the modified Bessel function of the first kind and order m. Furthermore, S was defined in (9.15) and

$$r_n \triangleq \frac{|\mathbf{a}_0^H \mathbf{a}_n|}{K}, \tag{9.37}$$

which may be interpreted as the relative sidelobe level.

Although the expression for P_n might appear somewhat unwieldy, it depends only on three characteristic quantities: the integrated SNR, S, the relative sidelobe level, r_n, and the number of snapshots, N.

A simpler expression is obtained if we consider the special case of a single snapshot. Equation (9.34) then simplifies to

$$P_n = Q\left(\sqrt{\frac{S}{2}\left(1-\sqrt{1-r_n^2}\right)}, \sqrt{\frac{S}{2}\left(1+\sqrt{1-r_n^2}\right)}\right) - \frac{1}{2} e^{-S/2} I_0\left(\frac{r_n S}{2}\right) \tag{9.38}$$

A fairly accurate approximation of this expression is given by [Ath02a]

$$P_n \approx \frac{1}{2} e^{-S/2} I_0\left(\frac{r_n S}{2}\right). \tag{9.39}$$

9.3.2 Stochastic signal model

In the stochastic model, both y_0 and y_n are zero mean. The second order moments are given by

$$\begin{aligned} \mathrm{E}\left[y_0 y_0^*\right] &= K^2\sigma_s^2 + K\sigma_n^2 && (9.40)\\ \mathrm{E}\left[y_n y_n^*\right] &= |\mathbf{a}_n^H\mathbf{a}_0|^2\sigma_s^2 + K\sigma_n^2 && (9.41)\\ \mathrm{E}\left[y_0 y_n^*\right] &= K\mathbf{a}_n^H\mathbf{a}_0\sigma_s^2 + \mathbf{a}_n^H\mathbf{a}_0\sigma_n^2. && (9.42) \end{aligned}$$

Results from [Pro95] can be used again to obtain P_n according to

$$P_n = \frac{1}{(1+q_n)^{2N-1}}\sum_{m=0}^{N-1}\binom{2N-1}{m}q_n^m, \tag{9.43}$$

where

$$q_n = \frac{\sqrt{1+\dfrac{4\sigma_n^2(K\sigma_s^2+\sigma_n^2)}{K^2\sigma_s^4(1-r_n^2)}}+1}{\sqrt{1+\dfrac{4\sigma_n^2(K\sigma_s^2+\sigma_n^2)}{K^2\sigma_s^4(1-r_n^2)}}-1}. \tag{9.44}$$

A simpler expression is obtained if we study the special case of a single snapshot:

$$P_n = \frac{1}{2}\left(1-\left(1+\frac{4\sigma_n^2(K\sigma_s^2+\sigma_n^2)}{K^2\sigma_s^4(1-r_n^2)}\right)^{-1/2}\right). \tag{9.45}$$

An interesting, and somewhat unexpected, property of the stochastic ML estimator appears if we study the behavior of P_n for high SNR. For $\sigma_s^2 \gg \sigma_n^2$, (9.43) reduces to

$$P_n = \binom{2N-1}{N-1}\cdot\frac{1}{\left[K(1-r_n^2)\,\mathrm{SNR}\right]^N}, \tag{9.46}$$

where $\mathrm{SNR} = \sigma_s^2/\sigma_n^2$. Hence, P_n decreases as $1/\mathrm{SNR}^N$ for sufficiently high SNR. For the case of a single snapshot, *i.e.* $N = 1$, (9.46) becomes

$$P_n = \frac{1}{K(1-r_n^2)\mathrm{SNR}}. \tag{9.47}$$

Furthermore, we make the approximation, $1-\sum_n P_n \approx 1$, which is very accurate for high SNR. Substituting into (9.26), the MSE approximation for high SNR becomes

$$\mathrm{E}\left[(\hat{u}-u_0)^2\right] \approx \ \mathrm{CRB} + \frac{1}{K\cdot\mathrm{SNR}}\sum_{n=1}^{N_p}\frac{(u_n-u_0)^2}{1-r_n^2}. \tag{9.48}$$

The stochastic CRB is proportional to 1/SNR at high SNR according to (9.17). Hence, the outlier terms in (9.48) cannot be neglected when letting $\text{SNR} \to \infty$. Therefore, the MSE will never reach the CRB when $\text{SNR} \to \infty$. The stochastic ML estimator is thus not asymptotically efficient as $\text{SNR} \to \infty$ for the case of a single snapshot. For $N > 1$, the outlier contribution to the MSE decreases as $1/\text{SNR}^N$ and can thus be neglected when letting $\text{SNR} \to \infty$. In this case, the efficiency of the ML estimator hinges upon whether the local errors are small enough to reach the CRB.

Somewhat surprisingly, we have found in simulations that not even the local errors appear to reach the CRB for high SNR. The distance between the ML MSE and the CRB seems to decrease inverse proportionally to the number of snapshots for high SNR. We have not verified this observation theoretically, nor have we found any other work that derives the high SNR asymptotics for the ML estimator under the stochastic signal model. We leave this as a topic for future research.

9.4 Simulation Results

In this section we will present some results of Monte Carlo simulations of the ML estimator under the deterministic and stochastic signal models. The results will be compared with the approximations that we described in the previous section. Our objectives are to justify the theoretical models we have developed and to gain some further insights into the threshold region behavior of the ML estimator.

The ML estimator was implemented by first computing the beamformer spectrum according to (9.7) on a coarse grid. Subsequently, a fine grid search was performed around each local maximum found in the first step. The final estimate was then obtained as the DOA corresponding to the maximum peak after the fine searches. The fine searches are needed for two reasons. Firstly, in order to locate the true maximum in the underlying continuous beamformer spectrum, the peaks must be sampled very close to their maxima. If the sampling is too coarse, there is a risk of picking the wrong peak if two peaks have similar magnitude. Secondly, we do not want the finite sampling grid to be the factor that bounds the variance of the estimator. Therefore, the sampling grid must be a fraction of the CRB if the error variance is ever going to reach the CRB.

9.4.1 Deterministic signal model

We will present the results for the deterministic case in a number of examples that illustrate various aspects of the threshold region behavior of the ML estimator. Since the threshold effect is most prominent for sparse arrays, we will use a particular sparse array structure called minimum redundancy arrays (MRAs) [Mof68] in the examples. In all examples we will use the 8 element MRA with element positions (normalized by $\lambda/2$) given by $\{0\ 1\ 2\ 11\ 15\ 18\ 21\ 23\}$. The modulus of the normalized beampattern, given by

$$|\bar{g}(u, u_0)| = |\mathbf{a}^H(u)\mathbf{a}(u_0)|/K, \tag{9.49}$$

is plotted in Figure 9.9, when steered to boresight (i.e. $u_0 = 0$). For comparison purpose we also plot the normalized beampattern of the 8 element ULA with $\lambda/2$ element spacing. The null-to-null beamwidth is 0.12 for the MRA and 0.50 for the ULA. The peak sidelobe level of the MRA is -5.7 dB whereas the ULA has its peak sidelobe at -13 dB.

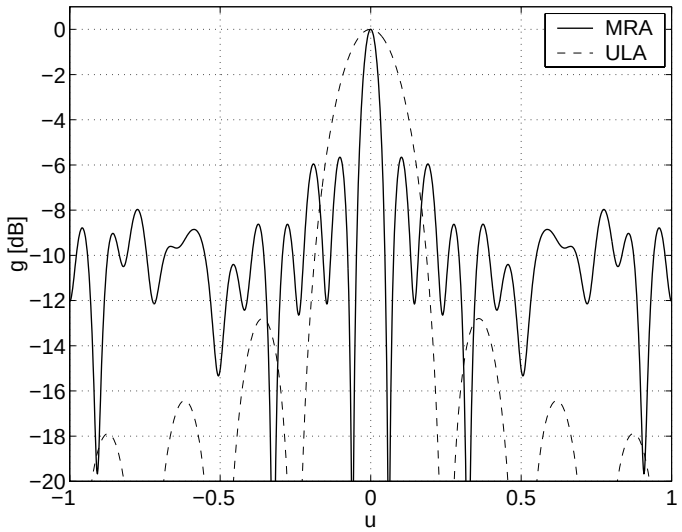

Figure 9.2: Beampattern for an 8 element MRA and ULA.

Example 9.1: Figure 9.3 shows a plot of the pairwise error probability, P_n, as a function of the integrated SNR, S, for the case of a single snapshot. The true DOA is $u_0 = 0$ and P_n is computed for $u_n = 0.10$, which corresponds to the sidelobe peak closest to the mainlobe in the beampattern of the 8 element MRA. The relative sidelobe level, r_n, is -5.7 dB. The solid curve shows P_n according to the exact expression in (9.34) and the dashed curve

the approximation in (9.39). The dots show the result of a Monte Carlo simulation with 10^7 trials for $S < 15$ dB and 10^8 trials for $S > 15$ dB. The agreement between the analytical expression in (9.34) and the simulation results is good, which is expected since no approximations were made in the derivation of this probability. We can also observe that the approximation in (9.39) is fairly accurate. □

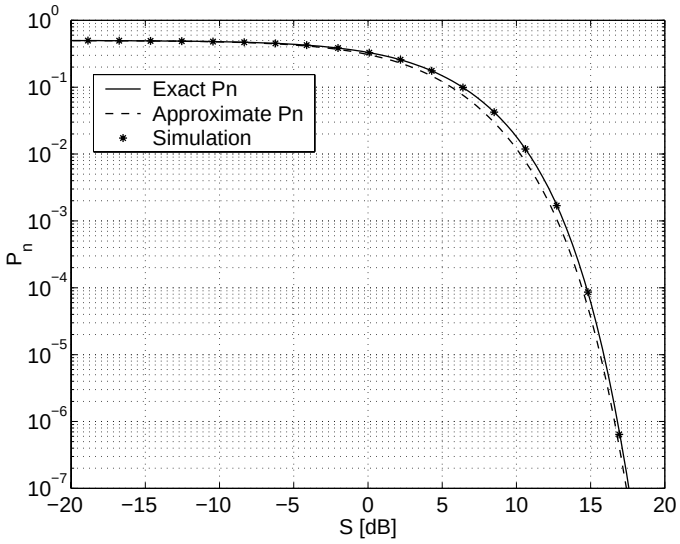

Figure 9.3: Pairwise error probability vs integrated SNR for the 8 element MRA; deterministic model, one snapshot.

Example 9.2: Figure 9.4 shows a plot of the pairwise error probability, P_n, as a function of the number of snapshots, N, a for constant signal, $s(t) = 0.5e^{j\pi/4}$, and noise variance $\sigma_n^2 = 1$. Otherwise, the same array and simulation parameters as in the previous example were used. The solid curve shows P_n according to (9.34) and the markers show the result of a Monte Carlo simulation with 10^7 trials. The figure shows that the analytical calculation of P_n is also correct for multiple snapshots. □

Example 9.3: In order to check the accuracy of the union bound and MSE approximations in (9.5) and (9.26), respectively, we performed a Monte Carlo simulation of the ML estimator. Figure 9.5 shows the probability of outlier and MSE for different values of the integrated SNR. We used the same array as in the other examples and we show the case of a single snapshot. The number of Monte Carlo trials was $2 \cdot 10^5$ for $S < 15$ dB and 10^6 for $S > 15$ dB. An outlier was defined as an estimate outside the first null on either side

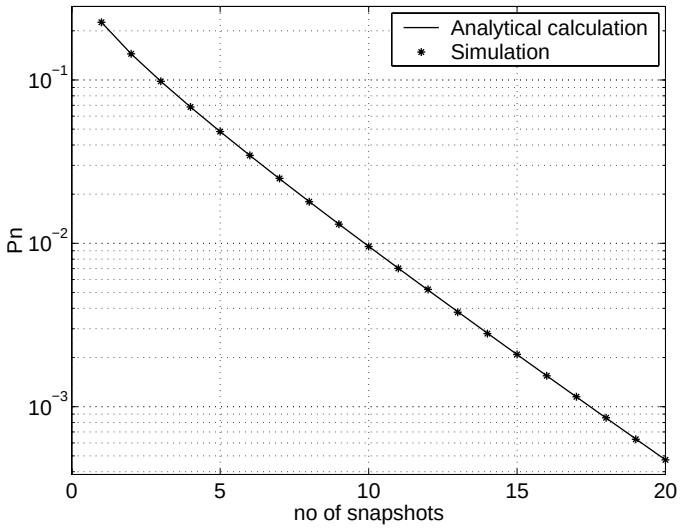

Figure 9.4: Pairwise error probability vs the number of snapshots for the 8 element MRA; deterministic model, $s(t) = 0.5e^{j\pi/4}, \sigma_n^2 = 1$.

of the mainlobe. We see that the union bound approximation is good for high SNR and poor for low SNR. Comparing with the MSE plot below we see that the union bound approximation is accurate in the threshold region.

The bottom plot in Figure 9.5 shows that the derived MSE approximation is very accurate in modeling the performance of the ML estimator. Only in the no information region does the approximation deviate from the simulation results. This is obviously due to that the union bound is not tight in this region. However, we do not consider this discrepancy as a deficiency of the MSE approximation, since the estimator is more or less useless anyway in this region.

We also compared the MSE approximation and simulation with the Cramér-Rao and Barankin bounds. The CRB was given in (9.14) and the Barankin bound was computed according to [TK99]. As expected, the CRB is attained for high SNR but does not predict the threshold effect. This is done by the Barankin bound, but the prediction of the SNR threshold is far too optimistic. □

Example 9.4: It is interesting to compare the MSE of an MRA with that of an ULA when both arrays have the same number of elements. Figure 9.6 shows the result from a simulation with the same 8 element MRA used before and an 8 element ULA with $\lambda/2$ element separation. The solid curves are the MSE approximations, the dashed are the corresponding CRBs, and the

markers are the results from Monte Carlo simulations with $2 \cdot 10^5$ trials. The number of snapshots in these simulations was one. As expected, the MRA has lower MSE for high SNR, but a poorer threshold performance. Note that the MSE of the MRA is an order of magnitude lower than that of the ULA for high SNR. This comes at the price of an increased SNR threshold by 3 dB (when defined as the point of departure from the corresponding CRB).□

Example 9.5: In this example, we illustrate how the different sidelobes contribute to the overall MSE. The n-th sidelobe contributes with $P_n(u_n - u_0)^2$ to the MSE approximation in (9.26). These terms are shown for all sidelobes of the 8 element MRA in the left plot in Figure 9.7. The dashed curve is the total MSE. The right plot shows the accumulated MSE as we include more and more sidelobes in the second sum in (9.26). The sidelobes have been ordered in magnitude. We can observe that the position of the threshold is to a great extent determined by the highest sidelobe. Including more sidelobes in the approximation moves the threshold slightly, but its main effect is to increase the MSE. □

Example 9.6: Finally, we show in Figure 9.8 how the pairwise error probability, P_n, depends on the relative sidelobe level, r_n. The solid curve shows P_n versus r_n according to the analytical expression in (9.34) and the markers show the result from a Monte Carlo simulation with 10^7 trials and one snapshot. The signal sample was $s = 1.8e^{j\pi/4}$, and the noise variance $\sigma^2 = 1$. □

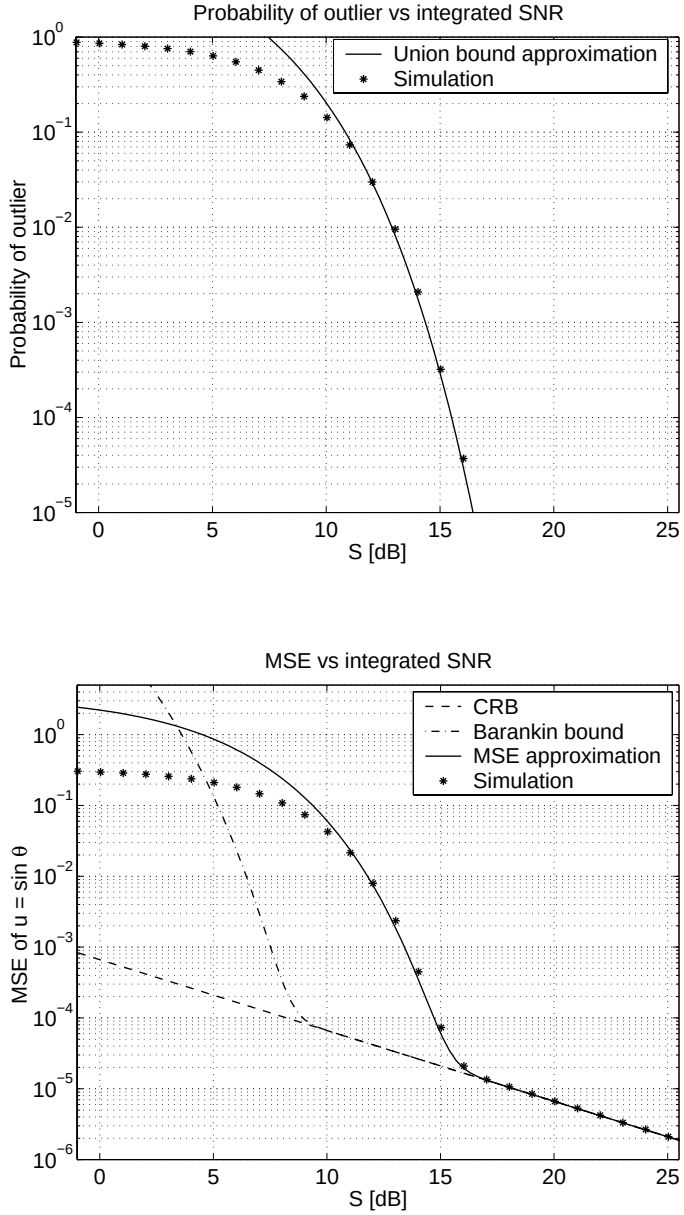

Figure 9.5: Monte Carlo simulation of the single source ML DOA estimator under the deterministic signal model. An 8 element MRA was employed and we show the result for a single snapshot. The simulation results are compared with the analytical calculations. The number of Monte Carlo trials was $2 \cdot 10^5$ for $S < 15$ dB and 10^6 for $S > 15$ dB. Top: The union bound approximation of the probability of outlier in (9.21) is compared with simulations for different values of the integrated SNR. Bottom: The MSE approximation in (9.26) is compared with simulations and the Barankin and Cramér-Rao bounds.

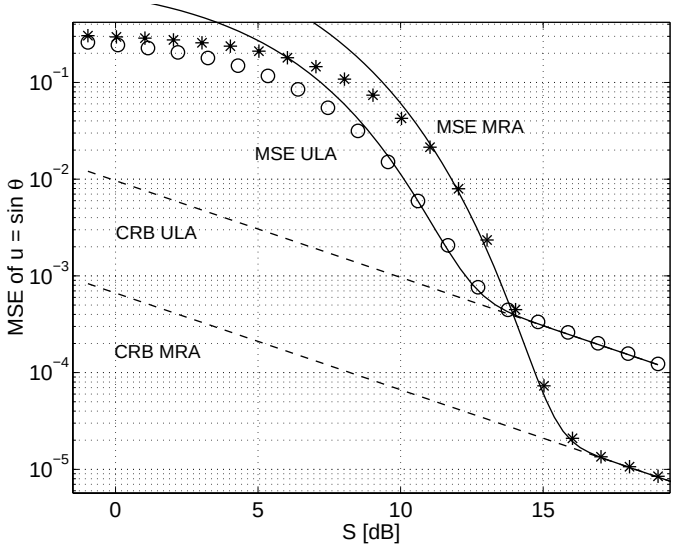

Figure 9.6: MSE vs integrated SNR for an 8 element MRA and ULA, respectively; deterministic model, 1 snapshot.

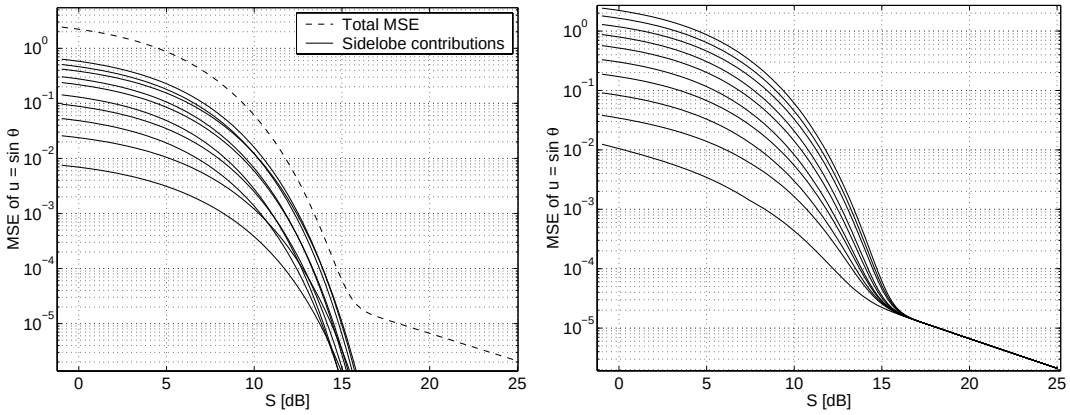

Figure 9.7: MSE contributions from the sidelobes of the 8 element MRA; deterministic model, 1 snapshot. Left: MSE contribution from each sidelobe (solid) and total MSE (dashed). Right: Accumulated MSE as more sidelobes are included in the second sum in (9.26).

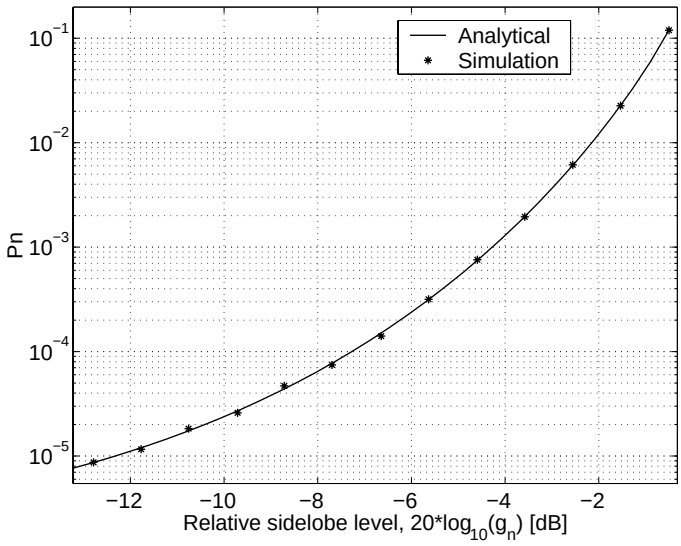

Figure 9.8: Pairwise error probability vs relative sidelobe level r_n; deterministic model, 1 snapshot, $s = 1.8e^{j\pi/4}$, $\sigma^2 = 1$.

9.4.2 Stochastic signal model

In this section we present some results of Monte Carlo simulations in order to support the analytical calculations for the stochastic signal model. This time, we will use a 4 element MRA with normalized element positions given by [0 1 4 6]. The normalized beampattern for this array is plotted in Figure 9.9 together with the 4 element ULA. The null-to-null beamwidth is 0.44 for the MRA and 1.0 for the ULA. The peak sidelobe level of the MRA is -5.3 dB whereas the ULA has its peak sidelobe at -11 dB.

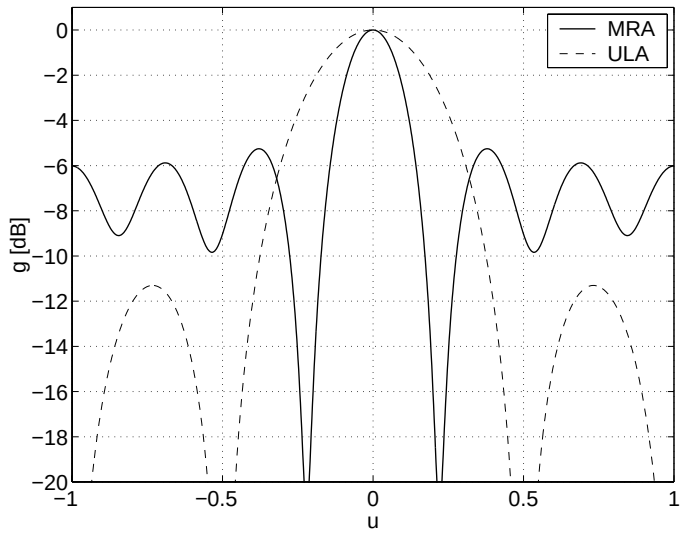

Figure 9.9: Beampattern for a 4 element MRA and ULA.

Figure 9.10 illustrates the accuracy of the union bound and MSE approximations for the stochastic signal model with 50 snapshots. The number of Monte Carlo trials was 10^6. Similar to the deterministic case, the approximations are accurate in the threshold and asymptotic regions, but poor in the no information region. In this case the Barankin bound also provides an overly optimistic prediction of the SNR threshold.

In order to check the accuracy of the approximations for fewer snapshots, we re-ran the previous simulation for 1, 2, 3, and 10 snapshots. The result is shown in Figure 9.11. For the case of one snapshot we can observe that the threshold effect is always present, even at very high SNR. Thus, as we have already shown theoretically, the estimator does not attain the CRB when the SNR is increased and we have only a single snapshot. The discrepancy between the simulations and the MSE approximation is due to that the union bound approximation is not so accurate for one snapshot under the stochastic model.

It is clear from the figure that the MSE and union bound approximations become more accurate when we increase the number of snapshots. This is natural since it is hard to characterize a stochastic signal with only a few samples. Even if we have a very high SNR, the outcome of the individual signal samples can be very small since the Gaussian PDF still has a considerable probability mass around zero.

We can also observe that the MSE does not quite reach the CRB for high SNR even when the number of snapshots is more than one. This effect is not caused by outliers, but is due to the local errors not being small enough to attain the CRB. Clearly, the offset from the CRB decreases when the number of snapshots increases.

In the final plot in Figure 9.12, we show the MSE versus the number of snapshots when the SNR is fixed at 3 dB. The number of Monte Carlo trials was 10^6. In this case we also get good agreement between the MSE approximation and the simulations.

As a final remark, we would like to comment on the bias of the estimators. The MSE can be divided into a variance and a bias part according to

$$\mathrm{MSE}(\hat{u}) \quad = \quad \mathrm{Var}(\hat{u}) + \mathrm{Bias}^2(\hat{u}). \tag{9.50}$$

The Cramér-Rao and Barankin bounds are bounds on the variance of any *unbiased* estimator. In the threshold region most estimators are biased. However, we have investigated the bias of the estimators in the Monte Carlo simulations and we found that in all cases the squared bias was negligible compared to the variance.

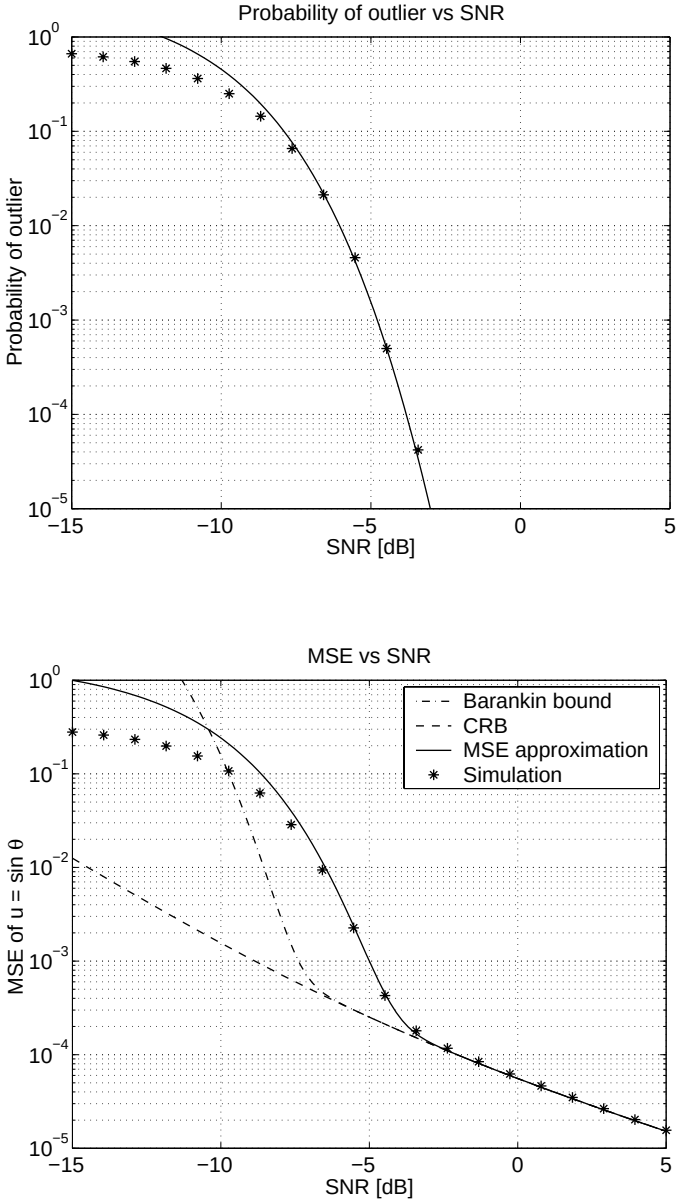

Figure 9.10: Monte Carlo simulation of the single source ML DOA estimator under the stochastic signal model. A 4 element MRA was employed and we show the result for 50 snapshots. The simulation results are compared with the analytical calculations. The number of Monte Carlo trials was 10^6. Top: The union bound approximation of the probability of outlier in (9.21) is compared with simulations for different values of the SNR. Bottom: The MSE approximation in (9.26) is compared with simulations and the Barankin and Cramér-Rao bounds.

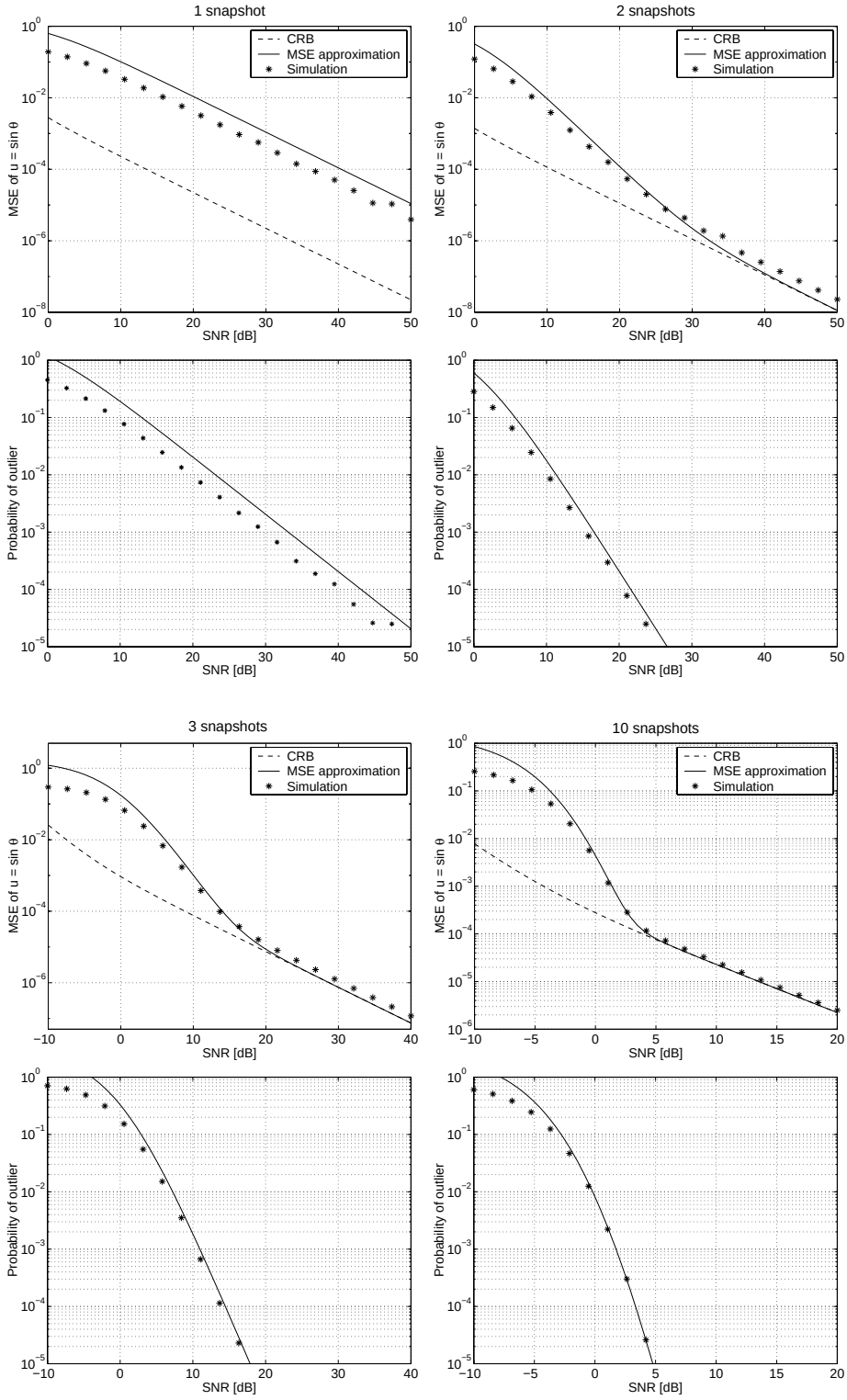

Figure 9.11: MSE and probability of outlier vs SNR, stochastic model for various numbers of snapshots.

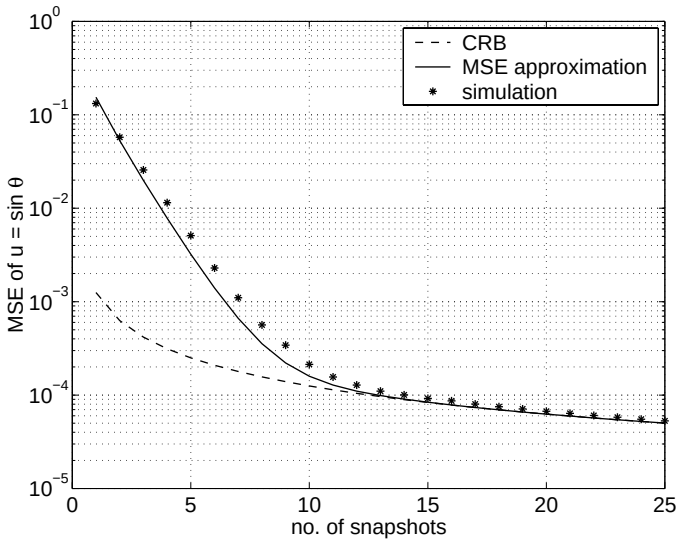

Figure 9.12: MSE vs number of snapshots, stochastic model, SNR = 3 dB.

Chapter 10

ML DOA Estimation of Multiple Sources

In this chapter we will generalize the analysis in the previous chapter to the case of multiple sources. The basic ideas behind the previous analysis carry over to the multi-source problem in a straightforward manner, but the technical details become more intricate.

10.1 Data Model and Problem Formulation

Generalizing the model in (9.1) to multiple sources is immediate:

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t), \qquad t = 1, \ldots, N, \tag{10.1}$$

where $\mathbf{A}(\boldsymbol{\theta}) = [\ \mathbf{a}(\theta_1) \ \ldots \ \mathbf{a}(\theta_M)\ ]$, $\mathbf{s}(t) = [\ s_1(t) \ \ldots \ s_M(t)\ ]^T$, and M is the number of sources. The zero subscripts on the true DOAs have been dropped for brevity. We assume that the number of sources is known, or has been estimated by some method. As in the single source problem we will study both the stochastic and deterministic models for the signals, $s_m(t)$.

It is convenient to express the model in (10.1) in matrix form according to

$$\mathbf{X} = \mathbf{A}(\boldsymbol{\theta})\mathbf{S} + \mathbf{N}, \tag{10.2}$$

where

$$\mathbf{X} = \left[\ \mathbf{x}(1) \ \ldots \ \mathbf{x}(N)\ \right] \tag{10.3}$$

$$\mathbf{S} = \left[\ \mathbf{s}(1) \ \ldots \ \mathbf{s}(N)\ \right] \tag{10.4}$$

$$\mathbf{N} = \left[\ \mathbf{n}(1) \ \ldots \ \mathbf{n}(N)\ \right]. \tag{10.5}$$

In this case we also consider the estimation of $u_k = \sin\theta_k, k = 1, \ldots, M$, for mathematical convenience.

10.2 Approximation of the MSE and Probability of Outlier

We will extend the MSE approximation for a single source to multiple sources simply by considering one dimension at a time in the multidimensional ML criterion function. First assume that the criterion function is parameterized by the DOAs only, i.e., there are no nuisance parameters that need to be considered. Such nuisance parameters could be, e.g., signal and noise powers. As we later will see, this assumption is valid for the deterministic ML criterion function, but not, in general, for the stochastic ML criterion.

Now, let $\bar{V}(\mathbf{u})$ denote the "asymptotic" ML criterion function. By asymptotic we mean high SNR or many snapshots, i.e., the noise should not have any effect on the positions of the peaks in the criterion function. We will return to more exact definitions of $\bar{V}(\mathbf{u})$ when we study the different signal models. Furthermore, let $\mathbf{u}_n = [u_{n1}, \ldots, u_{nM}]^T$ denote the n-th local maximum of $\bar{V}(\mathbf{u})$ and $\mathbf{u}_0 = [u_{01}, \ldots, u_{0M}]^T$ the global maximum corresponding to the true value of $\mathbf{u}$. We may also refer to these maxima as sidelobe and mainlobe peaks, respectively. To obtain an MSE approximation for the k-th DOA, we simply project the peaks in the multidimensional criterion function onto its k-th dimension. We thus obtain

$$\mathrm{E}\left[(\hat{u}_{0k} - u_{0k})^2\right] \approx \left(1 - \sum_{n=1}^{N_p} P_n\right) \mathrm{CRB} + \sum_{n=1}^{N_p} P_n \left(u_{nk} - u_{0k}\right)^2, \qquad (10.6)$$

where now

$$P_n \triangleq \Pr[V(\mathbf{u}_n) > V(\mathbf{u}_0)] \qquad (10.7)$$

is the multidimensional extension of the pairwise error probability in (9.22). Note that we have used same union bound approximation of the probability of outlier in (10.7) as in the single source case, i.e.

$$P_o \approx \sum_{n=1}^{N_p} P_n. \qquad (10.8)$$

In the following sections we will provide the details that are required for evaluating the approximations of the MSE and probability of outlier. In the

multi-source problem, the ML estimators are different under the deterministic and stochastic signal models. We will therefore treat these two cases in different sections. We start with the deterministic ML (DML) estimator and then present the results for the stochastic ML (SML) estimator. Results of Monte Carlo simulation will also be provided that corroborates the theoretical analysis.

10.3 Deterministic Maximum Likelihood

The ML estimator under the deterministic signal model is given by [Böh84, Wax85]

$$\hat{\mathbf{u}} = \arg\max_{\mathbf{u}} V(\mathbf{u}), \tag{10.9}$$

where

$$V(\mathbf{u}) = \mathrm{Tr}\Big\{\mathbf{\Pi}_{\mathbf{A}}(\mathbf{u})\hat{\mathbf{R}}\Big\}. \tag{10.10}$$

Here, $\mathbf{\Pi}_{\mathbf{A}}(\mathbf{u}) = \mathbf{A}(\mathbf{u})\left[\mathbf{A}^H(\mathbf{u})\mathbf{A}(\mathbf{u})\right]^{-1}\mathbf{A}^H(\mathbf{u})$ is the orthogonal projection matrix onto the column space of $\mathbf{A}$ and $\hat{\mathbf{R}} = \mathbf{X}\mathbf{X}^H/N$ is the sample covariance matrix. Under the deterministic signal model, we let the "asymptotic" ML criterion, $\bar{V}(\mathbf{u})$, be the ML criterion function in the absence of noise. From (10.2) and (10.10) the deterministic part of the ML criterion is given by

$$\begin{aligned}\bar{V}(\mathbf{u}) &= \frac{1}{N}\mathrm{Tr}\big\{\mathbf{\Pi}_{\mathbf{A}}(\mathbf{u})\mathbf{A}(\mathbf{u}_0)\mathbf{S}\mathbf{S}^H\mathbf{A}^H(\mathbf{u}_0)\big\} \\ &= \frac{1}{N}\big\|[\mathbf{A}(\mathbf{u})(\mathbf{A}^H(\mathbf{u})\mathbf{A}(\mathbf{u}))^{-1/2}]^H\mathbf{A}(\mathbf{u}_0)\mathbf{S}\big\|_F^2. \end{aligned}\tag{10.11}$$

We can interpret $\bar{V}(\mathbf{u})$ as a multidimensional generalization of the array beampattern. The columns of $\mathbf{A}(\mathbf{u})(\mathbf{A}^H(\mathbf{u})\mathbf{A}(\mathbf{u}))^{-1/2}$ contain M orthogonalized beams. According to (10.11), $\bar{V}(\mathbf{u})$ is obtained as the average output power from a matching operation between these beams and the received signal vectors in $\mathbf{A}(\mathbf{u}_0)\mathbf{S}$.

The CRB under the deterministic signal model is given by [SN89]

$$\mathrm{CRB}(\mathbf{u}) = \frac{\sigma^2}{2N}\left[\mathrm{Re}\big\{(\mathbf{D}^H\mathbf{\Pi}_{\mathbf{A}}^{\perp}(\mathbf{u})\mathbf{D}) \odot (\mathbf{S}\mathbf{S}^H)^T/N\big\}\right]^{-1}, \tag{10.12}$$

where $\mathbf{\Pi}_{\mathbf{A}}^{\perp}(\mathbf{u}) = \mathbf{I} - \mathbf{\Pi}_{\mathbf{A}}(\mathbf{u})$.

10.3.1 Pairwise Error Probability

The MSE approximation derived for the single source problem was easily generalized to multiple sources by considering one dimension at a time in the multidimensional criterion function. Now, the real challenge in analyzing the multi-source problem emerges when computing the pairwise error probabilities P_n. In this section, we will present a closed-form expression for the pairwise error probabilities given as an infinite series. Unfortunately, there are some practical problems associated with this expression. Therefore, we will subsequently present an accurate approximation to the desired probabilities that is more suitable for practical implementation.

Similar to the single source problem, we will show that finding the pairwise error probabilities amounts to computing the distribution of a quadratic form. In fact, anticipating the next section, the stochastic ML criterion can also be expressed as a quadratic form. Thus, we have a common framework that encompasses all cases considered in this part of the thesis.

To arrive at a quadratic form for the DML criterion function we use the following property of the trace operator [Lüt96]

$$\mathrm{Tr}\{\mathbf{ABCD}\} = \mathrm{vec}^T(\mathbf{D}^T)(\mathbf{C}^T \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{B}). \tag{10.13}$$

The DML criterion can then be written as

$$\begin{aligned} V(\mathbf{u}) &= \mathrm{Tr}\left\{\mathbf{\Pi_A}\hat{\mathbf{R}}\right\} = \frac{1}{N}\mathrm{Tr}\left\{\mathbf{\Pi_A}\mathbf{X}\mathbf{X}^H\right\} \\ &= \frac{1}{N}\,\mathrm{vec}^H(\mathbf{X})(\mathbf{I}_N \otimes \mathbf{\Pi_A})\,\mathrm{vec}(\mathbf{X}) = \frac{1}{N}\mathbf{y}^H\mathbf{\Omega}\mathbf{y} \end{aligned}$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix, $\mathbf{y} \triangleq \mathrm{vec}(\mathbf{X})$, and $\mathbf{\Omega} \triangleq \mathbf{I}_N \otimes \mathbf{\Pi_A}$. From (10.7) we can then express the pairwise error probability as

$$P_n = \Pr[\mathbf{y}^H\mathbf{\Psi}\mathbf{y} < 0], \tag{10.14}$$

where

$$\mathbf{\Psi} \triangleq \mathbf{\Omega}(\mathbf{u}_0) - \mathbf{\Omega}(\mathbf{u}_n) = \mathbf{I}_N \otimes \left[\mathbf{\Pi_A}(\mathbf{u}_0) - \mathbf{\Pi_A}(\mathbf{u}_n)\right]. \tag{10.15}$$

Thus, in order to compute P_n, we need to compute the distribution of the quadratic form $q = \mathbf{y}^H\mathbf{\Psi}\mathbf{y}$. It is a noncentral, indefinite quadratic form in complex Gaussian random variables. It is noncentral since $\mathrm{E}[\mathbf{y}] \neq \mathbf{0}$, and indefinite since q can be either positive or negative.

Much attention has been given in the literature to compute the distributions of various quadratic forms, see e.g. [JK70, MP92] and the references therein. Most results, however, are concerned with restricted cases, such

as positive definite, central quadratic forms or real-valued variables. Relatively little attention has been paid to the general case encountered here. In [PR96, Rap96, ML00] series expansions are derived for the distribution of noncentral, indefinite quadratic forms in complex Gaussian variables. We will use the result from [Rap96] to compute the pairwise error probability. However, before we present the result we need some preliminaries.

When computing the distribution of quadratic forms it is convenient to first reduce it to a diagonal form in independent variables with unit variance. To this end, let

$$\sigma^2\mathbf{\Psi} = \mathbf{U}\mathbf{D}\mathbf{U}^H \tag{10.16}$$

be the eigendecomposition of $\sigma^2\mathbf{\Psi}$. Here, $\mathbf{D}$ is a diagonal matrix with the eigenvalues on the diagonal and $\mathbf{U}$ contains the corresponding eigenvectors as columns. By making the transformation

$$\mathbf{w} = \frac{1}{\sigma}\mathbf{U}^H\mathbf{y}, \tag{10.17}$$

the quadratic form can be expressed as

$$z = \mathbf{w}^H\mathbf{D}\mathbf{w} = \sum_{k=1}^{KN} d_k|w_k|^2, \tag{10.18}$$

where d_k are the eigenvalues of $\sigma^2\mathbf{\Psi}$ and w_k is the corresponding element in $\mathbf{w}$. The moments of $\mathbf{w}$ are

$$\boldsymbol{\mu} \triangleq \mathrm{E}\left[\mathbf{w}\right] = \frac{1}{\sigma}\mathbf{U}^H \operatorname{vec}(\mathbf{A}\mathbf{S}) \tag{10.19}$$

$$\mathrm{E}\left[(\mathbf{w}-\boldsymbol{\mu})(\mathbf{w}-\boldsymbol{\mu})^H\right] = \mathbf{I}_{KN}. \tag{10.20}$$

Thus, the quadratic form has been written as a linear combination of noncentral independent χ^2 random variables. The number of degrees of freedom of the χ^2 variables is determined by the multiplicities of the eigenvalues d_k and the noncentrality parameters are determined by the elements in $\boldsymbol{\mu}$.

Next, we will show that it is possible to reduce the representation of the quadratic form in (10.18) to a function of the eigenelements of the single snapshot problem. To see this, let us first define

$$\mathbf{Q} = \mathbf{\Pi}_{\mathbf{A}}(\mathbf{u}_0) - \mathbf{\Pi}_{\mathbf{A}}(\mathbf{u}_n), \tag{10.21}$$

which is the matrix we would have in the quadratic form if we had only a single snapshot. Let the eigendecomposition of $\sigma^2\mathbf{Q}$ be given by

$$\sigma^2\mathbf{Q} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^H. \tag{10.22}$$

Assume that the nonzero eigenvalues of $\sigma^2\mathbf{Q}$ are distinct. We do not give a formal proof here that this is the case, but we simply state that in all the scenarios we have studied this has been a valid assumption. Accounting for possible multiple eigenvalues of $\sigma^2\mathbf{Q}$ is not difficult, but it complicates the notation.

Now, since $\mathbf{\Psi} = \mathbf{I}_N \otimes \mathbf{Q}$, there is a simple relationship between the eigendecompositions of $\mathbf{\Psi}$ and $\mathbf{Q}$. It is well known that if λ_A and λ_B are eigenvalues of $\mathbf{A}$ and $\mathbf{B}$, respectively, with associated eigenvectors $\mathbf{v}_A$ and $\mathbf{v}_B$; then $\lambda_A\lambda_B$ is an eigenvalue of $\mathbf{A} \otimes \mathbf{B}$ with eigenvector $\mathbf{v}_A \otimes \mathbf{v}_B$ [Lüt96]. Hence, the eigenvalues of $\mathbf{\Psi}$ will be same as those of $\mathbf{Q}$ with the exception that each eigenvalue will occur with multiplicity N. Therefore,

$$\mathbf{U} = \mathbf{I}_N \otimes \mathbf{E}, \qquad \mathbf{D} = \mathbf{I}_N \otimes \mathbf{\Lambda}. \tag{10.23}$$

The parameters that determine the quadratic form representation in (10.18) can now be expressed in terms of the eigendecomposition of $\sigma^2\mathbf{Q}$, which is a smaller matrix than $\sigma^2\mathbf{\Psi}$ if $N > 1$. The eigenvalues d_k are equal to λ_k, where λ_k are the eigenvalues of $\sigma^2\mathbf{Q}$, and they each have multiplicity N. Furthermore, the mean vector $\boldsymbol{\mu}$ in (10.19) can be written as

$$\boldsymbol{\mu} \;=\; \frac{1}{\sigma}\left(\mathbf{I}_N \otimes \mathbf{E}^H\right) \operatorname{vec}(\mathbf{AS}) = \frac{1}{\sigma}\operatorname{vec}(\mathbf{E}^H\mathbf{AS}). \tag{10.24}$$

Exact Expression

We are now ready to apply the result in [Rap96] to get the pairwise error probability.

Theorem 10.1 *Define* $\mathbf{Q} \triangleq \mathbf{\Pi}_{\mathbf{A}}(\mathbf{u}_0) - \mathbf{\Pi}_{\mathbf{A}}(\mathbf{u}_n)$ *and let the eigendecomposition of* $\sigma^2\mathbf{Q}$ *be given by*

$$\sigma^2\mathbf{Q} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^H, \tag{10.25}$$

where the eigenvectors and eigenvalues have been ordered so that $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_L\ 0 \ldots 0)$ *where* L *is the number of nonzero eigenvalues, which are assumed to be distinct. Furthermore, define*

$$\mu_j^2 \triangleq \frac{1}{\sigma^2}\sum_{n=1}^{N} |\mathbf{E}_j^H\mathbf{As}(n)|^2, \quad j = 1, \ldots, L, \tag{10.26}$$

where $\mathbf{E}_j$ *is the* j*-th column of* $\mathbf{E}$*. The pairwise error probability is then given*

by

$$\begin{aligned} P_n &= \Pr[V(\mathbf{u}_n) > V(\mathbf{u}_0)] = 1 - \exp\left[-\sum_{j=1}^{L} \mu_j^2\right] \\ &\times \sum_{\{k:\lambda_k>0\}} (-\lambda_k)^{-N} \sum_{m=N-1}^{\infty} \frac{1}{m!(m-N+1)!} \left(-\frac{\mu_k^2}{\lambda_k}\right)^{m-N+1} g_k^{(m)}. \end{aligned} \tag{10.27}$$

Here, $g_k^{(m)}$ *is given by*

$$g_k^{(m)} = \sum_{\ell=0}^{m-1} \binom{m-1}{\ell} g_k^{(\ell)} [\ln g_k]^{(m-\ell)}, \tag{10.28}$$

where

$$g_k^{(0)} = -\lambda_k \exp\left[\sum_{\substack{j=1 \\ j\neq k}}^{L} \frac{\mu_j^2}{\alpha_{kj}}\right] \prod_{j=1}^{L} \frac{1}{\alpha_{kj}^N}, \tag{10.29}$$

and

$$[\ln g_k]^{(m)} = \sum_{\substack{j=1 \\ j\neq k}}^{L} \frac{m!\lambda_j^m \mu_j^2}{\alpha_{kj}^{m+1}} + N \sum_{j=1}^{L} \frac{(m-1)!\lambda_j^m}{\alpha_{kj}^m}, \tag{10.30}$$

where $\alpha_{kj} = 1 - \lambda_j/\lambda_k$ *for* $k \neq j$ *and* $\alpha_{kk} \triangleq -1$.

Proof: See [Rap96].

Approximate Expression

The expression in Theorem 10.1 has some practical limitations. It takes a long time to compute the expression in (10.28) and, more seriously, we have also experienced numerical problems for several relevant cases. These are due to the series being composed of sums and differences of extremely large numbers being multiplied by extremely small numbers. The exponentials sometimes caused numbers in the series to become larger than the maximum floating point number representable on the computer. These problems were even more pronounced when we implemented the related result in [ML00].

Recognizing the numerical problems with the series expansion, a natural approach is to resort to some approximation that is more suitable for practical implementation. A common approach when computing the distribution of quadratic forms is to first compute the moment generating function analytically, and then use numerical integration for the inversion. Having reduced the quadratic form into its diagonal form as in the previous section, it is a simple matter to compute its moment generating function. The result is (see e.g. [Tur60, MP92] for details)

$$M(s) = \mathrm{E}\left[\exp(sq)\right] = \frac{\exp\left(\sum_{k=1}^{L} \frac{\mu_k^2 \lambda_k s}{1-\lambda_k s}\right)}{\prod_{k=1}^{L}(1-\lambda_k s)^N}, \tag{10.31}$$

where the same definitions as in Theorem 10.1 have been utilized. The probability density function (PDF) of q is then obtained by inverting the moment generating function using the Laplace inversion theorem. By integrating the PDF we can then obtain the pairwise error probability. Numerical procedures for this approach have been presented in [Ric80, Hel86, ML00, MLP02, Ath02b]. Although we found this approach efficient and numerically reliable, we will use another approach here that we have found even more efficient.

A very accurate technique used in statistics for approximating distributions is the so called *saddlepoint method* introduced by [Dan54]. For a review of saddlepoint methods in statistics, see [Rei88]. Here, we will use a form of saddlepoint approximation presented in [Kuo99]. The approximation is based on the cumulant generating function, which is the natural logarithm of the moment generating function. From (10.31) we obtain the cumulant generating function as

$$K(s) \;=\; \log M(s) = \sum_{k=1}^{L} \frac{\mu_k^2 \lambda_k s}{1-\lambda_k s} - N \sum_{k=1}^{L} \log(1-\lambda_k s). \tag{10.32}$$

To compute the saddlepoint approximation we also need the first two derivatives of the cumulant generating function. From (10.32) we obtain

$$K'(s) \;=\; \sum_{k=1}^{L} \frac{\mu_k^2 \lambda_k}{(1-\lambda_k s)^2} + N \sum_{k=1}^{L} \frac{\lambda_k}{1-\lambda_k s}, \tag{10.33}$$

$$K''(s) \;=\; \sum_{k=1}^{L} \frac{2\mu_k^2 \lambda_k^2}{(1-\lambda_k s)^3} + N \sum_{k=1}^{L} \frac{\lambda_k^2}{(1-\lambda_k s)^2}. \tag{10.34}$$

Now, the saddlepoint approximation of the cumulative distribution function (CDF) according to [Kuo99] is given by

$$F(x) \;=\; \Pr[q \leq x] = \Phi\left(w + \frac{1}{w}\log\left(\frac{v}{w}\right)\right) \tag{10.35}$$

where

$$\Phi(x) \;=\; \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt \tag{10.36}$$

$$w \;=\; \mathrm{sign}(s_0)\sqrt{2(s_0 x - K(s_0))} \tag{10.37}$$

$$v \;=\; s_0\sqrt{K''(s_0)}. \tag{10.38}$$

Here, s_0 is the so called saddlepoint satisfying the equation

$$K'(s_0) \;=\; x, \tag{10.39}$$

which can be numerically solved easily. It was shown in [Dan54] that this equation has a unique real root in the convergence region of the moment generating function. To ensure convergence of the moment generating function we pick the root in the interval

$$\frac{1}{\min(\lambda_k^-)} < s_0 < \frac{1}{\max(\lambda_k^+)}, \tag{10.40}$$

where λ_k^- and λ_k^+ denote the negative and positive eigenvalues, respectively. The pairwise error probability is then obtained from

$$P_n \;=\; F(0). \tag{10.41}$$

10.3.2 Simulations

In this section, we will present the results of some Monte Carlo simulations in order to support the theoretical analysis presented in the previous section. However, first let us briefly digress into a couple of delicate issues that are more important in this study than is usually the case.

Estimator Implementation

The DML estimator is obtained by maximizing the criterion function in (10.10) or, equivalently, by minimizing

$$\tilde{V}(\mathbf{u}) = \mathrm{Tr}\Big\{\mathbf{\Pi}_{\mathbf{A}}^{\perp}(\mathbf{u})\hat{\mathbf{R}}\Big\}, \tag{10.42}$$

which is more suitable for practical implementation. In the sequel we want to validate the theoretical model of the MSE and probability of outlier by means of Monte Carlo simulations. For such a validation to be of any value it is very important to ascertain that we find the actual global minimum and not a local minimum when implementing the DML estimator. Note that our aim here is to validate the theoretical models of the previous section and not to design a practical estimator for operation in a real system. In such a case other considerations can be important.

In order to ensure that we find the global minimum, the DML estimator is implemented by first performing a coarse grid search over the DML criterion function to find the approximate locations of all local minima. From every local minimum a Newton optimization is then performed to refine the initial local estimates. The final estimate is then obtained from the lowest value after the Newton searches. The Newton searches are needed, because if there are local minima of similar magnitude it is essential to locate the position of the minima accurately in order to choose the true global minimum.

In the local optimization we used a damped Newton algorithm, in which the estimate is iteratively calculated as

$$\mathbf{u}^{k+1} = \mathbf{u}^k - \mu_k \mathbf{H}^{-1}\mathbf{F}', \tag{10.43}$$

where $\mathbf{u}^k$ is the estimate at iteration k and μ_k is the step length. Furthermore, $\mathbf{H}$ is the Hessian matrix whose elements are given by

$$\{\mathbf{H}\}_{ij} = \frac{\partial^2 V(\mathbf{u})}{\partial u_i \partial u_j}, \tag{10.44}$$

and

$$\mathbf{F} = \left[\begin{array}{ccc} \frac{\partial V(\mathbf{u})}{\partial u_1} & \cdots & \frac{\partial V(\mathbf{u})}{\partial u_M} \end{array} \right]^T \tag{10.45}$$

is the gradient vector. For details on how to compute the Hessian and gradient, see [OVSN93]. The step length, μ_k, was chosen according to the scheme described in [OVSN93]. First we choose a $\mu < 1$ and take $\mu_k = \mu^i$ for the smallest integer $i \geq 0$ that causes an adequate decrease in the criterion function. The iterations in (10.43) continue until one of the following stopping criteria is fulfilled:

- $|\mathbf{H}^{-1}\mathbf{F}'|$ is less than a specified tolerance. This tolerance is chosen as some fraction of the CRB in order to obtain sufficient accuracy.
- No improvement can be found along the search direction (μ_k smaller than a tolerance).
- A maximum number of iterations has been reached.

Number of Required Monte Carlo Trials

The estimator implementation described in the previous section is computationally expensive since it involves a grid search followed by several Newton searches. In addition to this we want to model outliers, which occur with very low probability. Therefore, to obtain confidence in our simulation results, a large number of Monte Carlo trials is required. Consequently, validating the theoretical models of the previous section is very time-consuming. In order to minimize this time, we want to get a rough estimate of the number of Monte Carlo trials that are required for getting reliable results.

We confine ourselves here to a very simple analysis rather than a rigorous one. To get a simple analysis we make a cruder approximation of the MSE than the one in (10.6). Assume that for low SNR the estimates are uniformly distributed on $[-1,\ 1]$, resulting in a variance $= 2^2/12 = 1/3$. Assume further that $1 - \sum_n P_n \approx 1$. We then obtain a rough estimate of the MSE according to

$$\text{MSE} \approx \text{CRB} + \frac{1}{3} P_o \,. \tag{10.46}$$

Incidentally, this approximation also permits us to estimate the threshold in a simple manner. The threshold is roughly where the "small error MSE" equals the "outlier MSE", i.e. the threshold occurs approximately where $P_o = 3\,\text{CRB}$. This equation can be solved numerically to find for which values of SNR, number of snapshots, etc the threshold occurs.

We are now interested in estimating the MSE using a Monte Carlo simulation with a number of trials, say L. We estimate the MSE and probability of outlier from the Monte Carlo simulation according to

$$\widehat{\text{MSE}} = \frac{1}{L} \sum_{\ell=1}^{L} (\hat{u}_\ell - u_0)^2, \tag{10.47}$$

$$\widehat{P}_o = L_o / L, \tag{10.48}$$

where $\hat{u}_\ell$ is the DOA estimate obtained in the ℓ-th trial and L_o is the number of times $\hat{u}_\ell$ is outside the "mainlobe" in the asymptotic criterion function. The estimation error in $\widehat{\text{MSE}}$ is composed of two parts. The first part comes from the local errors and the second from the outliers. Since the outliers are rare events, this is the most difficult part to estimate with a limited number of trials. It is therefore reasonable to assume that the uncertainty in $\widehat{\text{MSE}}$ is only due to the uncertainty in $\widehat{P}_o$. In the analysis here, we therefore approximate $\widehat{\text{MSE}}$ by

$$\widehat{\text{MSE}} = \text{CRB} + \frac{1}{3} \widehat{P}_o \tag{10.49}$$

According to [Kay98, Appendix 2A], $\widehat{P}_o$ is approximately Gaussian for large number of trials with mean = P_o and variance = $P_o(1-P_o)/L$. The relative error $e = (\widehat{P}_o - P_o)/P_o$ then has the asymptotic PDF

$$e \in \mathrm{As}\mathcal{N}\left(0, \frac{1-P_o}{LP_o}\right). \tag{10.50}$$

Suppose we want $|e|$ to be less than ϵ with probability α. We then obtain

$$\epsilon \;=\; Q^{-1}\left(\frac{1-\alpha}{2}\right)\sqrt{\frac{1-P_o}{LP_o}}, \tag{10.51}$$

where Q^{-1} is the inverse function of

$$Q(x) \;=\; \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt. \tag{10.52}$$

We can now calculate a $100(1-\alpha)\%$ confidence interval for $\widehat{P}_o$ according to

$$P_o - \epsilon < \widehat{P}_o < P_o + \epsilon \tag{10.53}$$

where ϵ is given by (10.51).

In Figure 10.1 we plot the MSE (solid curve) according to (10.46) versus SNR and its corresponding 95% confidence interval (dashed curve) according to the analysis presented above. It is the same example for which we will later present the results of Monte Carlo simulations. The left plot is for 10^4 trials, which apparently gives rather poor confidence in the threshold region. The right plot is for 10^5 trials which should provide fairly accurate results also close to the threshold.

It should be noted that the number of Monte Carlo trials is most critical close to the threshold. From Figure 10.1 we can observe that the estimation error of the MSE is largest close to the threshold. For very low SNR, the estimate of P_o is accurate even with a moderate number of trials since the probability of outlier is fairly high in this region. For very high SNR, the estimate of P_o is very poor since this probability is extremely low in this region. However, since P_o is then extremely low, it has almost no influence on the MSE and it does not matter how accurately we estimate it. In the threshold region, P_o is low, but high enough to affect the MSE. Therefore, more Monte Carlo trials are needed in the threshold region.

A Monte Carlo simulation with the DML estimator described previously took about 14 hours for each SNR value when using 10^5 trials and implemented in Matlab on a 1 GHz computer. In contrast, computing the MSE

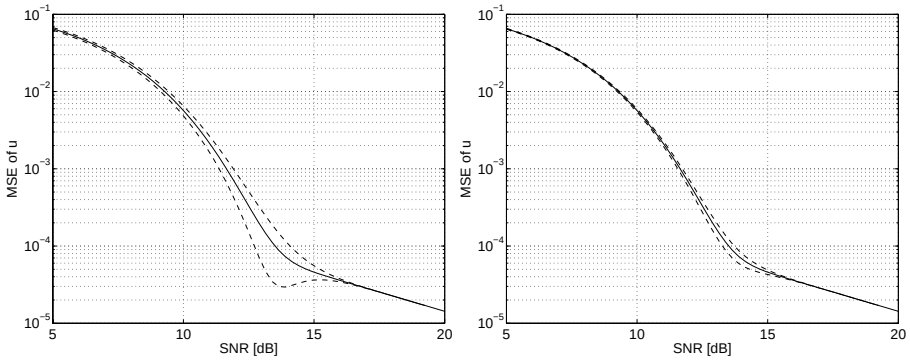

Figure 10.1: MSE (solid curve) and 95% confidence interval (dashed curves) for estimated MSE using Monte Carlo simulations with 10^4 trials (left plot) and 10^5 trials (right plot), respectively.

approximation using the saddlepoint method described in the previous section is done in a split second. This demonstrates a clear advantage with having a model of the threshold region performance rather than having to perform time-consuming Monte Carlo simulations.

Simulation Results

We will now present the results of some Monte Carlo simulations of the threshold region performance of the DML estimator and compare it to the performance predictions obtained from the theoretical models that we have just derived. As in the single source case, we will evaluate the models with sparse arrays since the threshold effect is most salient for such arrays.

First, we would like to validate the analytical calculation of the pairwise error probability, P_n. An example with an 8 element MRA with element positions, in units of $\lambda/2$, given by $\{0\ 1\ 2\ 11\ 15\ 18\ 21\ 23\}$ and two sources at $\mathbf{u}_0 = [0.2\ 0.4]$ was implemented. The two source signals were constant and equal, i.e. $s_1(n) = s_2(n) = s$. The number of snapshots was $N = 5$.

Figure 10.2 shows the results of a simulation of the pairwise error probability, P_n, versus SNR $= |s|^2/\sigma^2$ for $\mathbf{u}_n = [0.20\ -0.37]$, which is the position of the highest "sidelobe peak" in the asymptotic ML criterion function. The dots show the result of a Monte Carlo simulation with 10^7 trials and the solid curve is the result obtained using the saddlepoint approximation of P_n. Clearly, the agreement between the simulation results and the theoretical analysis is satisfactory.

The next example illustrates the accuracy of the approximations of the probability of outlier and MSE. Figure 10.3 shows the results from an example

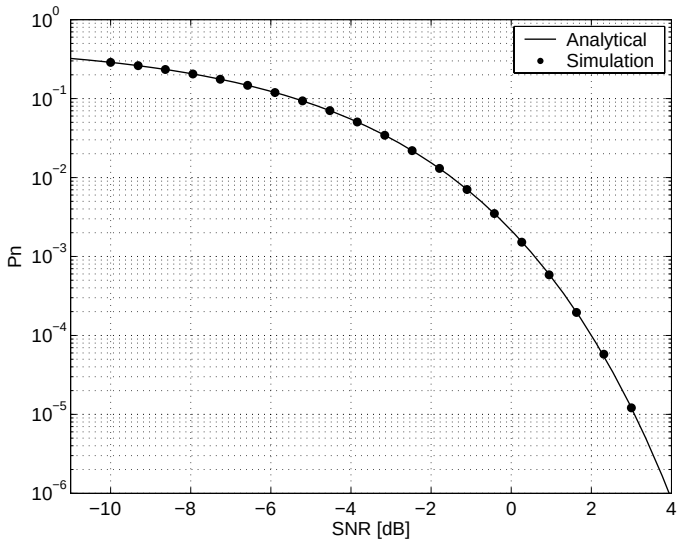

Figure 10.2: P_n for $\mathbf{u}_n = [0.20 \quad -0.37]$ vs SNR for an example with two sources at $\mathbf{u}_0 = [0.2\ 0.4]$ and an 8 element MRA; deterministic signal model, 5 snapshots, 10^7 Monte Carlo trials.

with a 4 element MRA with element positions $\{0\ 1\ 4\ 6\}$ and two sources at $\mathbf{u}_0 = [0\ 0.2]$. Thus, according to Section 9.4.2, the source separation is slightly less than half the null-to-null beamwidth of this array. Again, the two source signals were constant and equal, i.e. $s_1(n) = s_2(n) = s$, but this time with 7 snapshots. The number of Monte Carlo trials was 10^5.

The top plot in Figure 10.3 shows the probability of outlier versus SNR per space-time sample, obtained from the simulations and union bound approximation in (10.8), respectively. The estimates of probability of outlier from the simulations were obtained by counting the number of estimates outside the "mainlobe" region of the ML criterion function, and dividing by the total number of trials.

The bottom plot shows the MSE versus SNR per space-time sample for the source at $u = 0$ obtained from the simulation and equation (10.6) respectively. The DML estimator was implemented according to Section 10.3.2. The dashed line is the CRB from [SN89], which does not indicate any threshold effect. We did not implement the Barankin bound for the multi-source problem, since this bound was not a good predictor of the ML performance for the single source case, and there is no reason to expect that it would be better for the multi-source case. The figure clearly shows that our theoretical models of the threshold region performance under the deterministic signal model are accurate also for the multi-source problem.

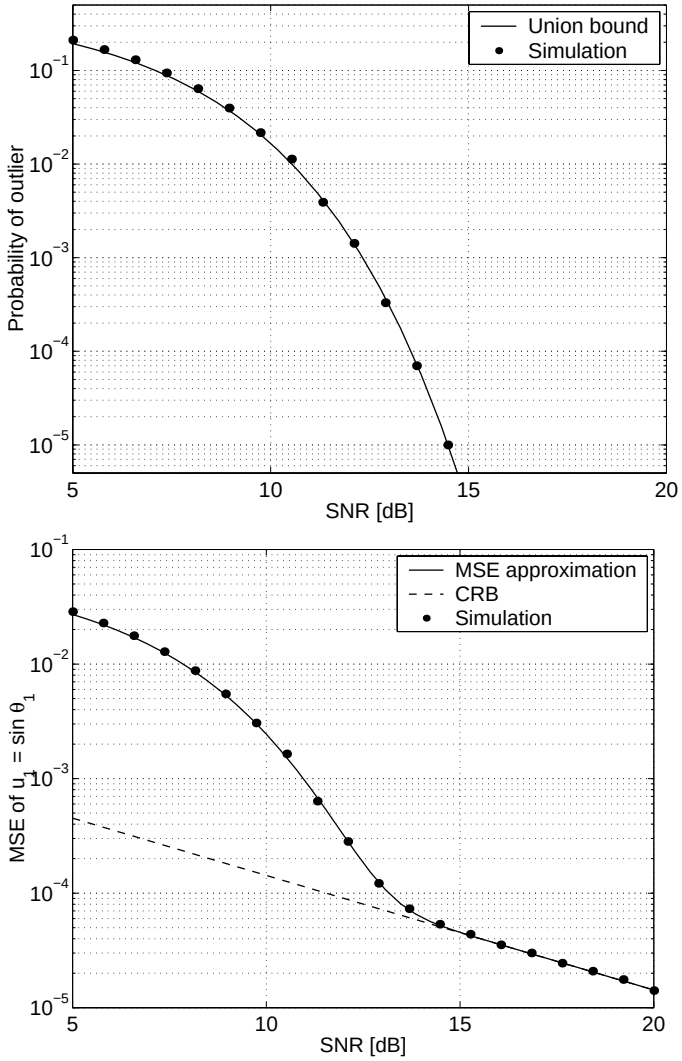

Figure 10.3: Monte Carlo simulation of the multi-source ML DOA estimator under the deterministic signal model. A 4 element MRA was employed and there were two sources at $\mathbf{u}_0 = [0\ 0.2]$. The number of snapshots was 7. The simulation results are compared with the analytical calculations. The number of Monte Carlo trials was 10^5. Top: The union bound approximation of the probability of outlier in (10.8) is compared with simulations for different values of the SNR. Bottom: The MSE approximation in (10.6) is compared with simulations and the Cramér-Rao bound.

10.4 Stochastic Maximum Likelihood

Under the stochastic signal model, the ML criterion function to be minimized is given by (see Section 1.4.2)

$$V(\mathbf{u}, \mathbf{P}, \sigma^2) \;=\; \log|\mathbf{R}| + \mathrm{Tr}\Big\{\mathbf{R}^{-1}\hat{\mathbf{R}}\Big\}, \tag{10.54}$$

where

$$\mathbf{R}(\mathbf{u}, \mathbf{P}, \sigma^2) \;=\; \mathbf{A}(\mathbf{u})\mathbf{P}\mathbf{A}^H(\mathbf{u}) + \sigma^2\mathbf{I}. \tag{10.55}$$

A complication with the criterion function in (10.54) is that it depends on the nuisance parameters in $\mathbf{P}$ and σ^2. This makes the optimization problem considerably more difficult. However, as stated in [Böh86] and proved later in [SN95] the criterion function can be concentrated with respect to $\mathbf{P}$ and σ^2. According to the above cited references, the ML estimates of $\mathbf{P}$ and σ^2 are obtained by inserting the ML estimates of $\mathbf{u}$ in the following expressions:

$$\hat{\mathbf{P}}(\mathbf{u}) \;=\; \mathbf{A}^{\dagger}(\mathbf{u})\left[\hat{\mathbf{R}} - \hat{\sigma}^2(\mathbf{u})\mathbf{I}\right]\mathbf{A}^{\dagger H}(\mathbf{u}), \tag{10.56}$$

$$\hat{\sigma}^2(\mathbf{u}) \;=\; \frac{1}{K-M}\mathrm{Tr}\Big\{\mathbf{P}^{\perp}_{\mathbf{A}}(\mathbf{u})\hat{\mathbf{R}}\Big\}. \tag{10.57}$$

The concentrated form of the SML criterion function is then obtained by substituting these expressions back into (10.54). This gives

$$V(\mathbf{u}) = \log\left|\mathbf{A}(\mathbf{u})\hat{\mathbf{P}}(\mathbf{u})\mathbf{A}^H(\mathbf{u}) + \hat{\sigma}^2(\mathbf{u})\mathbf{I}\right|. \tag{10.58}$$

Unfortunately, there is a problem with this concentrated form of the criterion function. The estimate of the signal covariance matrix in (10.56) was derived by minimizing the unconcentrated criterion function with respect to $\mathbf{P}$ over the set of Hermitian matrices. Therefore, there is no guarantee that the resulting estimate is positive semi-definite, as it should be. Whenever $\hat{\mathbf{P}}$ in (10.56) is positive semi-definite, the DOA estimates based on the unconcentrated and concentrated criterion functions, respectively, coincide [Bre88]. However, if $\hat{\mathbf{P}}$ is indefinite, they may differ.

In the case of a non-singular signal covariance matrix $\mathbf{P}$ (i.e. non-coherent sources) this is no problem when the number of snapshots is large. The reason for this is that $\hat{\mathbf{P}}$ is a consistent estimate of $\mathbf{P}$ as $N \to \infty$, which means that $\hat{\mathbf{P}}$ will be positive definite for sufficiently large N. In the case of coherent sources, the approach of [SOVM96] may be adopted. Therein, an ML estimator based on a square-root parameterization of the signal covariance

matrix subject to a rank constraint was derived. However, for large N, the difference between this and the ordinary ML estimator was found to be small.

Apparently, there seems to be no problem with indefinite signal covariance matrix estimates in the asymptotic regime. Unfortunately, this is not the case in the threshold region. As we later will see, when $\hat{\mathbf{P}}$ is indefinite it is very likely that the DOA estimate will be an outlier. Conversely, if an outlier has occurred, it is very likely that it was based on an indefinite $\hat{\mathbf{P}}$. Therefore, the estimator that minimizes the concentrated SML criterion function in (10.58) will have significantly poorer threshold region performance than the one that minimizes the unconcentrated criterion function in (10.54).

The concentrated criterion is easier to minimize than the unconcentrated one, since it depends on fewer parameters. On the other hand, as we just mentioned, it has poorer threshold region performance. Furthermore, it seems more difficult to analyze the statistical distribution of the concentrated criterion than the unconcentrated one due its complicated dependence on the measured data. As we will show later, the unconcentrated criterion can be expressed as a quadratic form and, hence, we can use an analysis similar to the one that we used for the deterministic case.

Our initial idea was to derive approximations of the MSE and probability of outlier for the unconcentrated criterion and then to validate the theory with Monte Carlo simulations using the concentrated criterion. Evidently, this is not a viable approach. We can then identify two approaches to follow:

- Accept the poorer threshold region performance of the estimator based on the concentrated criterion, and develop the theory for this case. Unfortunately, we were not able to find the pairwise error probabilities that are associated with this criterion.

- Develop the theory for the estimator based on the unconcentrated criterion and use the same estimator in the Monte Carlo simulations. Unfortunately, this approach also presents problems. Although we are able to develop the theory, the estimator in its general form is too complex to use in a sufficient number of Monte Carlo trials for a reliable validation.

Due to the these difficulties we are not able provide a complete picture of the threshold region performance of the SML estimator. We must be content with an analysis of a couple of special cases. These are:

1. Unconcentrated criterion function with the assumption that $\mathbf{P}$ and σ^2 are known.

2. Unconcentrated criterion with unknown $\mathbf{P}$ and σ^2, but the sources are known to be uncorrelated, i.e. $\mathbf{P}$ is diagonal.

The first case is not particularly realistic in radar applications, whereas the second one is. However, we will find that the performances in these cases are very similar. We will also compare these cases with the case of the concentrated criterion through simulations. A discussion on the difficulties with the case of unconcentrated criterion and correlated sources will also be given. All in all, we dare to say that this section will at least provide a decent understanding of the threshold region performance of the SML estimator.

10.4.1 Pairwise Error Probability

Similarly to the case of the DML estimator, we will first express the ML criterion function as a quadratic form. Let $\boldsymbol{\eta}$ be the parameter vector that contains the unknown parameters. In the case of known $\mathbf{P}$ and σ^2, this is simply $\boldsymbol{\eta} = \mathbf{u}$. In the case of unknown, but diagonal, $\mathbf{P}$ and σ^2, it is given by $\boldsymbol{\eta} = [\ \mathbf{u}^T \quad P_1 \quad \cdots \quad P_M \quad \sigma^2\ ]^T$, where $P_1, \ldots, P_M$ are the signal powers. From (10.54) the probability that the criterion function at a point $\boldsymbol{\eta}_n$ is lower than the criterion function at the true value $\boldsymbol{\eta}_0$ can be expressed as

$$\begin{aligned} P_n &\triangleq \Pr[V(\boldsymbol{\eta}_n) < V(\boldsymbol{\eta}_0)] \\ &= \Pr\left[\mathrm{Tr}\Big\{\left[\mathbf{R}_n^{-1} - \mathbf{R}_0^{-1}\right]\hat{\mathbf{R}}\Big\} < \log\left|\mathbf{R}_n^{-1}\mathbf{R}_0\right|\right], \end{aligned} \tag{10.59}$$

where $\mathbf{R}_0 \triangleq \mathbf{R}(\boldsymbol{\eta}_0)$ and $\mathbf{R}_n \triangleq \mathbf{R}(\boldsymbol{\eta}_n)$. Defining $\mathbf{Q} \triangleq \mathbf{R}_n^{-1} - \mathbf{R}_0^{-1}$ this can be written as

$$P_n = \Pr\left[\mathrm{Tr}\big\{\mathbf{Q}\mathbf{X}\mathbf{X}^H\big\} < N\log\left|\mathbf{R}_n^{-1}\mathbf{R}_0\right|\right]. \tag{10.60}$$

Using (10.13) we obtain

$$\begin{aligned} P_n &= \Pr\left[\mathrm{vec}^H(\mathbf{X})(\mathbf{I}_N \otimes \mathbf{Q})\,\mathrm{vec}(\mathbf{X}) < N\log\left|\mathbf{R}_n^{-1}\mathbf{R}_0\right|\right] \\ &= \Pr\left[\mathbf{y}^H\boldsymbol{\Psi}\mathbf{y} < N\log\left|\mathbf{R}_n^{-1}\mathbf{R}_0\right|\right], \end{aligned} \tag{10.61}$$

where $\boldsymbol{\Psi} \triangleq \mathbf{I}_N \otimes \mathbf{Q}$ and $\mathbf{y} \triangleq \mathrm{vec}(\mathbf{X})$. Similarly to the DML case we have arrived at an expression involving a quadratic form in complex Gaussian random variables.

The quadratic form obtained in the SML case has some properties that are different from the DML case. In the SML case the quadratic form is central since $\mathrm{E}[\mathbf{y}] = \mathbf{0}$. This will allow the PDF of the quadratic form to be expressed as a finite sum as compared to the infinite sum in (10.27) for

the DML case. The reason for this is that a nonzero mean of the quadratic form implies essential singularities in its moment generating function. The PDF of the quadratic form is given by the inverse Laplace transform of the moment generating function. This can be obtained, e.g., by residue calculus. Due to the essential singularities, the Laurent series expansion will be an infinite series. In the zero-mean case, however, there are only finite-order poles leading to a finite sum expression for the residues. Another property that differs from the DML case is that the covariance of $\mathbf{y}$ is not the identity matrix, but is given by $\mathrm{E}\left[\mathbf{y}\mathbf{y}^H\right] = \mathbf{I}_N \otimes \mathbf{R}_0$. In both the DML and SML cases the quadratic form is indefinite.

Exact Probability

First we will present a closed form expression for the pairwise error probability, P_n, given as a finite sum. In some cases, this expression is not very practical. Therefore, an accurate approximation of P_n is given in the next section.

Theorem 10.2 *Let $d_i, i = 1, \ldots, m$, be the distinct eigenvalues of $\mathbf{R}_0\mathbf{R}_n^{-1}$ that satisfy $d_i \neq 1$ and denote the corresponding multiplicities by n_i. Let $\gamma_i \triangleq (d_i - 1)/2$ be ordered such that $\gamma_1 > \ldots > \gamma_p > 0 > \gamma_{p+1} > \ldots > \gamma_m$. Assume first that $|\mathbf{R}_0| < |\mathbf{R}_n|$ and define*

$$\lambda_i = \begin{cases} \gamma_i, & i = 1, \ldots, p \\ -\gamma_i, & i = p+1, \ldots, m \end{cases} \tag{10.62}$$

$$c_i = \frac{\lambda_i^{-2Nn_i + N\sum_{k=1}^{m} n_k}}{(Nn_i - 1)!\, 2^{Nn_i} \Big[\prod_{k=1}^{p} (\lambda_i + \lambda_k)^{Nn_k}\Big] \Big[\prod_{\substack{k=p+1 \\ k\neq i}}^{m} (\lambda_i - \lambda_k)^{Nn_k}\Big]} \tag{10.63}$$

Furthermore, define

$$B_i^{(\beta)} = \sum_{\beta'=0}^{\beta-1} \binom{\beta-1}{\beta'} A_i^{(\beta-1-\beta')} \sum_{\beta''=0}^{\beta'-1} \binom{\beta'-1}{\beta''} A_i^{(\beta'-1-\beta'')} \sum_{\beta'''=0}^{\beta''-1} \cdots\ A_i^{(0)} \tag{10.64}$$

where

$$\begin{aligned} A_i^{(v)} &= v! \sum_{\substack{k=p+1 \\ k\neq i}}^{m} \frac{Nn_k(2\lambda_k)^{v+1}\lambda_i^{v+1}}{(\lambda_i-\lambda_k)^{v+1}} \\ &+ (-1)^{v+1} v! \sum_{k=1}^{p} \frac{Nn_k(2\lambda_k)^{v+1}\lambda_i^{v+1}}{(\lambda_i+\lambda_k)^{v+1}}, \quad v \geq 0. \end{aligned} \tag{10.65}$$

Then, the pairwise error probability is given by

$$P_n = F(N \log |\mathbf{R}_0 \mathbf{R}_n^{-1}|), \tag{10.66}$$

where

$$\begin{aligned} F(x) &= \sum_{i=p+1}^{m} c_i \sum_{\beta=0}^{Nn_i-1} \binom{Nn_i-1}{\beta} (-1)^{Nn_i-1-\beta} (2\lambda_i)^{Nn_i-\beta} (Nn_i-1-\beta)! \\ &\times B_i^{(\beta)} e^{x/2\lambda_i} \sum_{k=0}^{Nn_i-1-\beta} \frac{(-x)^k}{k!(2\lambda_i)^k}. \end{aligned} \tag{10.67}$$

If $|\mathbf{R}_n| < |\mathbf{R}_0|$*, then change the definition of* γ_i *to* $\gamma_i \triangleq (1-d_i)/2$*. The pairwise error probability is in this case given by*

$$P_n = 1 - F(-N \log |\mathbf{R}_0 \mathbf{R}_n^{-1}|). \tag{10.68}$$

Proof: See Appendix 10A

Approximate Probability

Although the pairwise error probability as given in Theorem 10.2 is in closed form, this expression has some practical limitations. When the number of snapshots is small the expression is fast to compute and also numerically reliable. However, for a large number of snapshots the recursion in (10.64) takes a long time to compute, and the final expression is also sensitive to round-off errors since it involves differences between very large numbers. For the examples to be presented later, the expression in Theorem 10.2 works fine up to approximately 12 snapshots.

For larger numbers of snapshots, one must resort to some approximation that is more practical to compute. A fast and accurate approximation based on numerical integration was given in [Imh61]. However, since we found the saddlepoint method, described previously in Section 10.3.1 for the DML

problem, to be even faster and more accurate, we will adopt this approach here.

In order to apply the saddlepoint method we need the moment generating function of the quadratic form

$$q = \mathbf{y}^H \mathbf{\Psi} \mathbf{y}. \tag{10.69}$$

Defining

$$\mathbf{\Sigma} = \mathrm{E}\left[\mathbf{y}\mathbf{y}^H\right] = \mathbf{I}_N \otimes \mathbf{R}_0, \tag{10.70}$$

we obtain the moment generating function as [MP92]

$$M(s) = \mathrm{E}\left[\exp(sq)\right] = \frac{1}{\prod_{k=1}^{L}(1-\lambda_k s)^{r_k}}, \tag{10.71}$$

where $\lambda_k, k = 1, \ldots, L$, are the distinct eigenvalues of $\mathbf{\Sigma\Psi}$ and r_k is the multiplicity of λ_k. The cumulant generating function is given by

$$K(s) = \log M(s) = -\sum_{k=1}^{L} r_k \log(1-\lambda_k s), \tag{10.72}$$

and its first two derivatives are easily found as

$$K'(s) = \sum_{k=1}^{L} \frac{r_k \lambda_k}{1-\lambda_k s}, \tag{10.73}$$

$$K''(s) = \sum_{k=1}^{L} \frac{r_k \lambda_k^2}{(1-\lambda_k s)^2}. \tag{10.74}$$

The saddlepoint approximation of the CDF is then given by

$$\tilde{F}(x) = \Phi\left(w + \frac{1}{w}\log\left(\frac{v}{w}\right)\right), \tag{10.75}$$

where

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \tag{10.76}$$

$$w = \mathrm{sign}(s_0)\sqrt{2(s_0 x - K(s_0))}, \tag{10.77}$$

$$v = s_0 \sqrt{K''(s_0)}. \tag{10.78}$$

The saddlepoint s_0 is obtained by numerically finding a solution to

$$K'(s_0) \quad = \quad x \tag{10.79}$$

in the interval

$$\frac{1}{\min(\lambda_k^-)} < s_0 < \frac{1}{\max(\lambda_k^+)}, \tag{10.80}$$

where λ_k^- and λ_k^+ denote the negative and positive eigenvalues, respectively. The pairwise error probability is then obtained from

$$P_n \quad = \quad \tilde{F}(N \log \left|\mathbf{R}_n^{-1}\mathbf{R}_0\right|). \tag{10.81}$$

10.4.2 MSE Approximation

In this section we will discuss how to apply the MSE approximation to the SML estimation problem. The first issue is the definition of the "asymptotic" criterion function, $\bar{V}(\boldsymbol{\eta})$, that should be used in the MSE approximation in (10.6). We cannot use the noise-free criterion function as we did in the DML case, since this would make $\mathbf{R}$ singular. Instead, we will replace $\hat{\mathbf{R}}$ in (10.54) with

$$\mathbf{R}_0 \quad = \quad \mathbf{A}(\mathbf{u}_0)\mathbf{P}_0\mathbf{A}^H(\mathbf{u}_0) + \sigma_0^2\mathbf{I}, \tag{10.82}$$

where the zero subscripts indicate the true parameter values. With this definition, $\bar{V}(\boldsymbol{\eta})$, is equal to the limiting criterion function as $N \to \infty$.

The next issues are which CRB to use and how to deal with the nuisance parameters in $\mathbf{P}$ and σ^2 when applying the MSE approximation to the DOA estimates. These issues depend on the assumptions we make, and we discuss one case at a time in the following paragraphs.

Unconcentrated criterion, known P and σ^2

In this case, the only unknown parameters are the DOAs in $\mathbf{u}$. Therefore, the problem is similar to the DML case and we can use the same MSE approximation. The only thing that differs from the DML case are the pairwise error probabilities and the CRB. The pairwise error probabilities are obtained from Theorem 10.2, and we take the CRB from [Tre02, Section 8.4.3], reproduced below for convenience

$$\begin{aligned} \mathrm{CRB}(\mathbf{u}) \quad &= \quad \frac{1}{2N}\Big[\mathrm{Re}\Big\{(\mathbf{P}\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A}\mathbf{P}) \odot \big(\mathbf{D}^H\mathbf{R}^{-1}\mathbf{D}\big)^T \\ &+ \quad (\mathbf{P}\mathbf{A}^H\mathbf{R}^{-1}\mathbf{D}) \odot \big(\mathbf{P}\mathbf{A}^H\mathbf{R}^{-1}\mathbf{D}\big)^T \Big\}\Big]^{-1}. \end{aligned} \tag{10.83}$$

Unconcentrated criterion, uncorrelated signals, unknown P and σ^2

In this case we need to deal with the nuisance parameters in $\mathbf{P}$ and σ^2. The SML criterion function is a multidimensional function with a multimodal structure in the DOA hyperplane. The multimodality is due to the highly nonlinear dependence of the criterion function on the DOA parameters. However, $\mathbf{P}$ and σ^2 enter the criterion function in a nicer way than the DOAs, since $\mathbf{R}$ is linear in $\mathbf{P}$ and σ^2. Actually, we have found by numerical investigation of the SML criterion function that the multimodal structure is limited to the DOA hyperplane only. Furthermore, in the case of uncorrelated signals, we have found that the positions of the DOA mainlobe and sidelobe peaks in the asymptotic criterion remain approximately the same when varying $\mathbf{P}$ and σ^2.

In order to support this claim, we will show some plots of the unconcentrated SML criterion function for an example with two sources. The unknown parameters are in this case $\boldsymbol{\eta} = [\ u_1 \quad u_2 \quad P_1 \quad P_2 \quad \sigma^2\ ]^T$. It is difficult to visualize a function in such high dimensions. Therefore, we show in Figure 10.4 different slices of the high-dimensional unconcentrated SML criterion function. The different plots show $V(\boldsymbol{\eta})$ when varying two parameters while keeping the others fixed. The top left plot shows $V(P_1, P_2)$ for fixed u_1, u_2, and σ^2. Clearly, there is no modal structure in these dimensions. The top right plot shows $V(P_1, u_1)$ for fixed u_2, P_2, and σ^2. We see the familiar mainlobe/sidelobe structure in the DOA dimension, but no such behavior in the signal power dimension. Furthermore, we observe that the mainlobe/sidelobe structure remains the same when varying the signal power. The bottom left plot shows $V(u_1, \sigma^2)$ for fixed u_2, P_1, and P_2. The observations made for the previous plot hold also for this one. Finally, the bottom right plot shows $V(P_1, \sigma^2)$ for fixed u_1, u_2, and P_2. This surface has no ambiguity structure.

The observations that the mainlobe/sidelobe structure is limited to the DOA dimensions, and that this structure is fairly unaffected by the values of the other parameters simplify matters to a great extent. We then need only to consider the DOA sidelobe structure when using the MSE approximation. Therefore, we can use the same MSE approximation as in the previous cases. We evaluate P_n for $\boldsymbol{\eta}_n = [\ \mathbf{u}_n \quad \mathbf{P}_0 \quad \sigma_0^2\ ]^T$, where $\mathbf{P}_0$ and σ_0^2 are the true values of $\mathbf{P}$ and σ^2. Furthermore, we need to use the CRB for uncorrelated signals which is given by [Tre02, Section 8.4.2]

$$\mathrm{CRB}(\mathbf{u}) \quad = \quad \left[\mathbf{J}_{uu} - \mathbf{J}_{u\alpha}\mathbf{J}_{\alpha\alpha}^{-1}\mathbf{J}_{u\alpha}^T\right]^{-1}, \tag{10.84}$$

where

$$\begin{aligned}
\mathbf{J}_{uu} &= 2N\,\mathrm{Re}\Big\{(\mathbf{P}\mathbf{A}^H\mathbf{R}^{-1}\mathbf{D})\odot\big(\mathbf{P}\mathbf{A}^H\mathbf{R}^{-1}\mathbf{D}\big)^T \\
&+ (\mathbf{P}\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A}\mathbf{P})\odot\big(\mathbf{D}^H\mathbf{R}^{-1}\mathbf{D}\big)^T\Big\}, && (10.85)\\
\mathbf{J}_{u\alpha} &= \begin{bmatrix}\mathbf{J}_{us} & \mathbf{J}_{un}\end{bmatrix}, && (10.86)\\
\mathbf{J}_{\alpha\alpha} &= \begin{bmatrix}\mathbf{J}_{ss} & \mathbf{J}_{sn}^T\\ \mathbf{J}_{sn} & \mathbf{J}_{nn}\end{bmatrix}, && (10.87)\\
\mathbf{J}_{us} &= 2N\,\mathrm{Re}\Big\{(\mathbf{P}\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A})\odot\big(\mathbf{A}^H\mathbf{R}^{-1}\mathbf{D}\big)^T\Big\}, && (10.88)\\
\mathbf{J}_{un} &= 2N\,\mathrm{Re}\big\{\mathrm{Diag}\{\mathbf{P}\mathbf{A}^H\mathbf{R}^{-2}\mathbf{D}\}\big\}, && (10.89)\\
\mathbf{J}_{ss} &= N\Big[(\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A})\odot\big(\mathbf{A}^H\mathbf{R}^{-1}\mathbf{A}\big)^T\Big], && (10.90)\\
\mathbf{J}_{sn} &= N\,\mathrm{Diag}\{\mathbf{A}^H\mathbf{R}^{-2}\mathbf{A}\}, && (10.91)\\
\mathbf{J}_{nn} &= N\,\mathrm{Tr}\{\mathbf{R}^{-2}\}. && (10.92)
\end{aligned}$$

Unconcentrated criterion, correlated signals, unknown P and σ^2

This is the most difficult case, since when the signals are correlated the positions of the DOA sidelobe peaks will change with the signal correlation. The full dimension of the SML criterion must then be taken into account when applying the MSE approximation. The pairwise error probabilities are still given by Theorem 10.2, but care has to be taken at which points the parameters should be evaluated at. Although the MSE approximation would probably work also for this case, it would be very difficult to validate such a model with Monte Carlo simulations with the computing power available today. It would require on the order of 10^5 Monte Carlo trials[1], where each trial would involve a search for the global minimum in a high-dimensional function with a complicated multimodal structure.

Concentrated criterion

This case is equivalent to first case in the sense that the criterion function depends only the DOA parameters. Therefore we can use the MSE approximation used before. However, as alluded to previously, the minimizer of the concentrated criterion is not necessarily the same as the minimizer of the unconcentrated criterion. In fact, we will later see that when using the concentrated criterion the threshold will be significantly higher than when using

[1]This figure is of course scenario dependent. In the examples that will be presented later in the simulations section, 10^5 trials are required for a reasonably reliable validation. See further Section 10.3.2.

the unconcentrated one. This behavior could be predicted if we computed the pairwise error probabilities for the concentrated criterion. Unfortunately, this appears to be a difficult task and we leave that as an open research problem. Nevertheless, it is interesting to compare the threshold region performance of this estimator with the one based on the unconcentrated criterion. Therefore, we will also include this case in the simulations.

10.4.3 Simulation Results

In this section we will present the results of some Monte Carlo simulations in order to illustrate how accurately we can model the threshold region performance of the SML estimator. As we have seen, some different cases appear depending on how we treat the nuisance parameters in $\mathbf{P}$ and σ^2. Of course, this will have an impact also on the implementation of the estimator. In the following, we will present simulation results for all the cases discussed above except for the case of unconcentrated criterion, correlated signals, unknown $\mathbf{P}$ and σ^2. In all cases the estimates are obtained using first a coarse grid search followed by a Newton search from each local minimum as described in Section 10.3.2.

Let us describe the standard example we have used in all simulations that we will present in the sequel. The same four element minimum redundancy array is used as in the DML examples. We consider a scenario with two sources having DOAs $\mathbf{u}_0 = [-0.1\ 0.1]$. The source signals have equal powers and are uncorrelated. The SNR is defined as $\text{SNR} = P/\sigma^2$, where P is the signal power. In the plots of the MSE we will compare the results of Monte Carlo simulations with the MSE approximation and the CRB.

Pairwise Error Probability

Before we show results from Monte Carlo simulations of the MSE, we would like to first check the analytical calculation of the pairwise error probabilities in Section 10.4.1. Figure 10.5 shows the result from a Monte Carlo simulation of P_n versus SNR for a case where $\mathbf{u}_0 = [-0.1\ 0.1]$ and $\mathbf{u}_n = [-0.015\ 0.67]$. This corresponds to the pairwise error probability for the global minimum and the lowest local minimum in the asymptotic likelihood function when $\mathbf{P}$ and σ^2 are known. The pairwise error probabilities were evaluated at the true values of $\mathbf{P}$ and σ^2. The number of trials was 10^7 and the number of snapshots was 5. We can observe that there is a perfect match between the simulation and the analytical calculation. No difference between the exact expression in Theorem 10.2 and the saddlepoint approximation could be observed.

Unconcentrated criterion, known P and σ^2

Figure 10.6 shows the MSE obtained from our approximation and a Monte Carlo simulation with 10^5 trials. The CRB is also included for comparison. The MSE is plotted versus the SNR and the number of snapshots was 12. We see that the MSE approximation is not as accurate as in the DML case. However, the MSE approximation predicts the SNR threshold very accurately. The discrepancy between the MSE approximation and the simulation result is due to that the MSE does not attain the CRB for high SNR. If we knew the high SNR variance of the SML estimator, the MSE approximation would be very accurate. As usual the MSE approximation is poor in the no information region due to the union bound approximation.

In Figure 10.7 we show a scatter plot of all 10^5 estimates for SNR = -1.6 dB, which is in the threshold region. A contour plot of the criterion function is also drawn. We see that most estimates are in the mainlobe region, and that the outliers are clustered around the sidelobe peaks. A color 3-D view of this plot is printed on the cover of this thesis.

In Figure 10.8 we show the result when varying the number of snapshots, keeping the SNR fixed at 5 dB. The number of Monte Carlo trials was 10^5. Also in this case we get a good prediction of the threshold, but the MSE approximation is not as tight as in the DML case. The MSE has not quite reached the CRB for 25 snapshots, but from the figure it appears likely that it would if the number of snapshots was increased further.

Unconcentrated criterion, uncorrelated signals, unknown P and σ^2

In Figure 10.9 we show the result when $\mathbf{P}$ and σ^2 are unknown but the signals are known to be uncorrelated. The number of snapshots was 12 and Monte Carlo trials 10^5. We observe from the figure that the result is very similar to the previous case.

In this example, the criterion function to be minimized is 5-dimensional. Repeating a grid search in the full dimension and Newton searches from all local minima 10^5 times is a formidable task for most computers of today. However, as we showed previously, the DOA sidelobe structure remains approximately the same when the other parameters are varied. Therefore, we fixed the values of P_1, P_2, and σ^2 and performed the grid search over u_1 and u_2 only. We used the sample variances of the actual outcomes of $\mathbf{s}(t)$ and $\mathbf{n}(t)$ as the fixed values of P_1, P_2, and σ^2 when the DOA grid search was performed. Of course, these values are not known in any practical situation, but our purpose here is to model the fundamental performance limitations of the ML estimator and not to design a practical implementation. Newton-searches

from all local minima were then performed in the full 5-D space.

Concentrated criterion

We now turn to the case of the concentrated criterion. The result is shown in Figure 10.10. The number of snapshots was 12 and Monte Carlo trials 10^5. In the MSE approximation we used the probabilities in Theorem 10.2 and the CRB for unstructured (besides Hermitian) $\mathbf{P}$ given in [SN90b]. In this case, the MSE approximation does not provide an accurate model for the threshold behavior of the concentrated SML estimator. As alluded to previously, this is due to that minimizing the unconcentrated and concentrated criteria is equivalent only if the estimate of the signal covariance matrix, $\hat{\mathbf{P}}$, is positive semi-definite. Apparently, this cannot be the case here since the threshold region performance is significantly poorer when using the concentrated criterion.

We kept track of all indefinite $\hat{\mathbf{P}}$s and two relevant statistics are displayed in Figure 10.11. The left plot shows the percentage outliers among those DOA estimates that were based on an indefinite $\hat{\mathbf{P}}$. An outlier was defined as an estimate outside the mainlobe area. Clearly, almost every time an indefinite $\hat{\mathbf{P}}$ was obtained, it resulted in an outlier when estimating the DOAs. The right plot shows the percentage indefinite estimates among those outcomes that resulted in a DOA outlier. We observe that about 90% of all outliers were based on an indefinite $\hat{\mathbf{P}}$. Hence, we conclude that there is a strong connection between the threshold behavior when using the concentrated criterion and the definiteness of $\hat{\mathbf{P}}$.

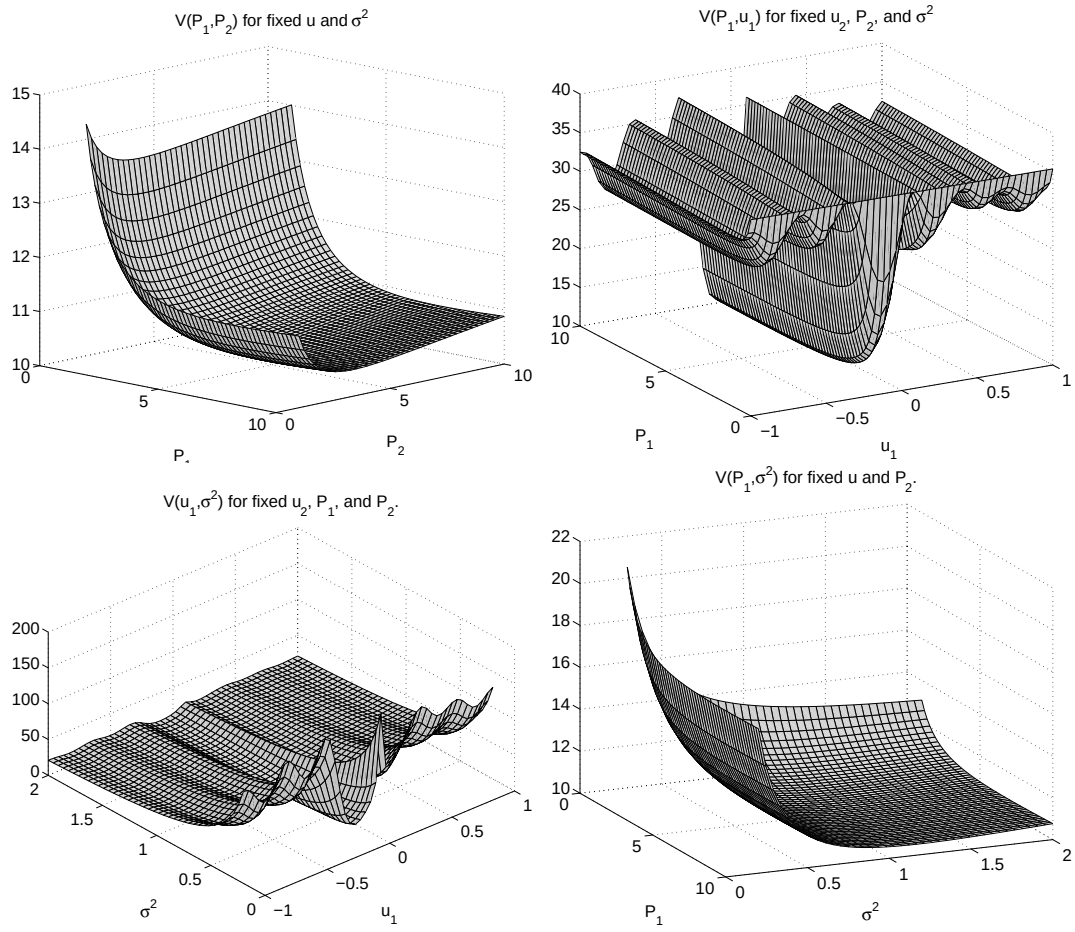

Figure 10.4: Different slices of the 5-dimensional unconcentrated stochastic ML criterion function for an example with two sources at $\mathbf{u} = [-0.1\ 0.1]$ and a 4 element MRA; $P_1 = P_2 = 5, \sigma^2 = 1, N = 20$. Top left: V as a function of P_1 and P_2 for $\mathbf{u} = [-0.1\ 0.1], \sigma^2 = 1$. Top right: V as a function of P_1 and u_1 for $u_2 = 0.1, P_2 = 5, \sigma^2 = 1$. Bottom left: V as a function of σ^2 and u_1 for $u_2 = 0.1, P_1 = P_2 = 5$. Bottom right: V as a function of P_1 and σ^2 for $\mathbf{u} = [-0.1\ 0.1], P_2 = 5$.

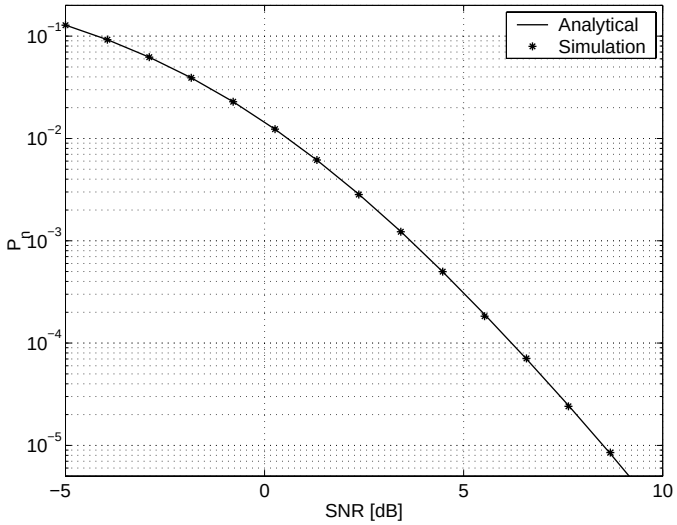

Figure 10.5: Pairwise error probability vs SNR per space-time sample for an example with two sources and a 4 element MRA. $\mathbf{u}_0 = [-0.1\ 0.1], \mathbf{u}_n = [-0.015\ 0.67], N = 5, \sigma^2 = 1, \text{SNR} = P$

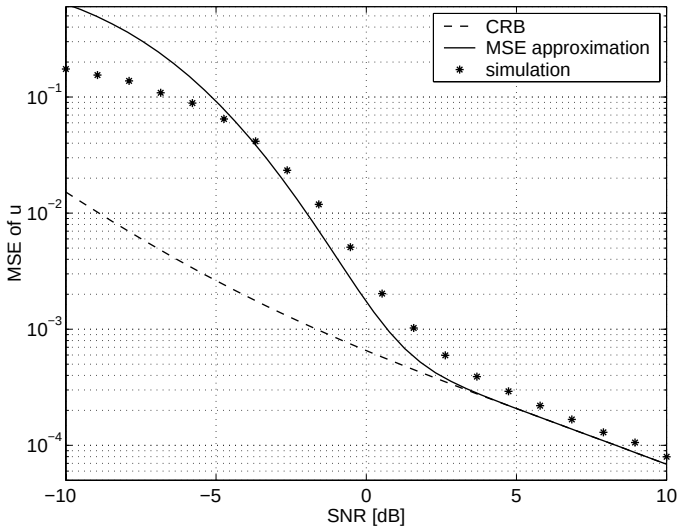

Figure 10.6: MSE vs SNR per space-time sample for an example with two sources and a 4 element MRA; unconcentrated criterion, known $\mathbf{P}$ and σ^2, 12 snapshots.

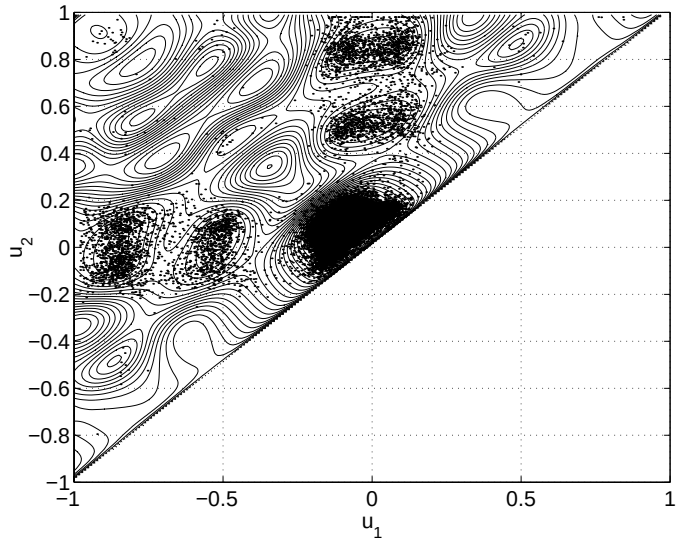

Figure 10.7: The SML criterion function and scatter plot of 10^5 estimates for an example with two sources and a 4 element MRA; unconcentrated criterion, known $\mathbf{P}$ and σ^2, 12 snapshots, SNR = -1.6 dB.

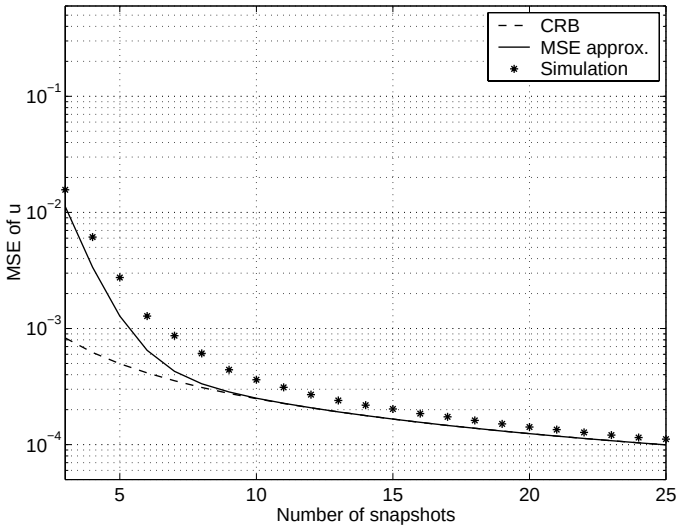

Figure 10.8: MSE vs number of snapshots per space-time sample for an example with two sources and a 4 element MRA; unconcentrated criterion, known $\mathbf{P}$ and σ^2, SNR = 5 dB.

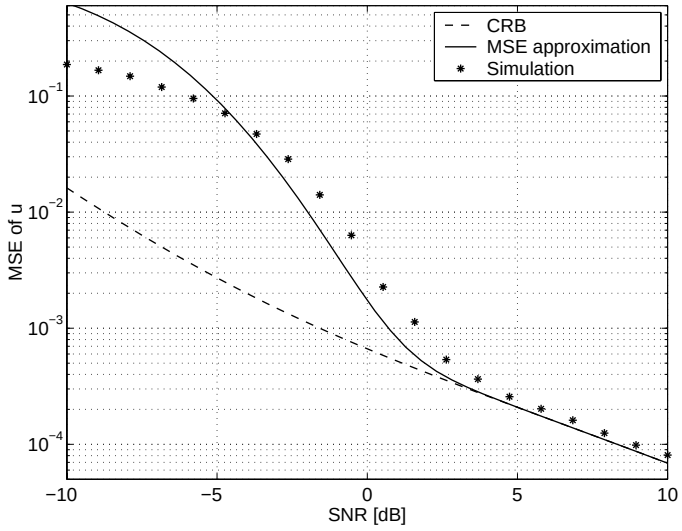

Figure 10.9: MSE vs SNR per space-time sample for an example with two sources and a 4 element MRA; unconcentrated criterion, uncorrelated signals, unknown $\mathbf{P}$ and σ^2, 12 snapshots.

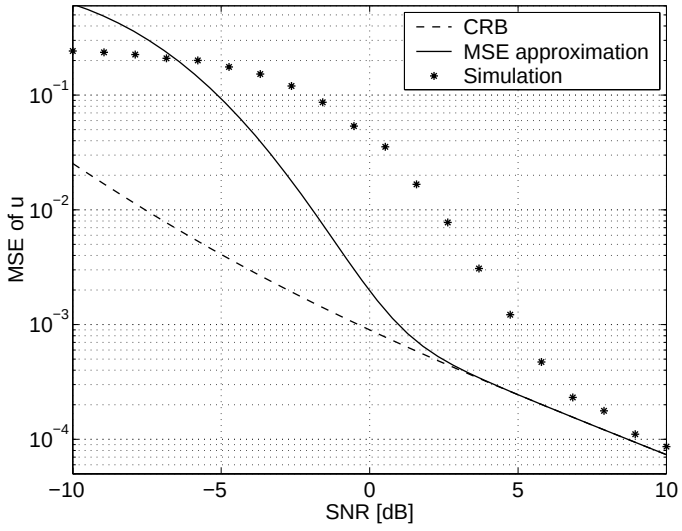

Figure 10.10: MSE vs SNR per space-time sample for an example with two sources and a 4 element MRA; concentrated criterion, 12 snapshots.

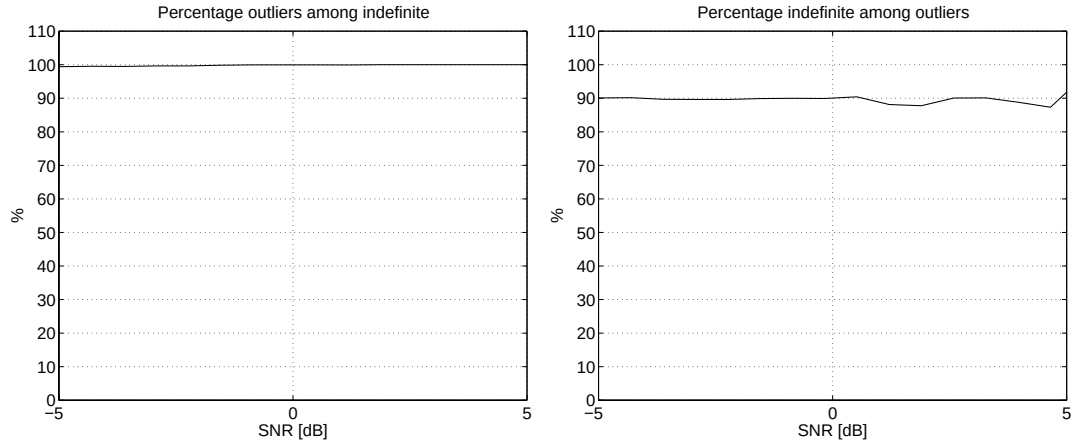

Figure 10.11: Percentage outliers among indefinite estimates of the signal covariance matrix and percentage indefinite signal covariance matrix estimates among the outliers. Concentrated SML estimator.

10A Proof of Theorem 10.2

To compute the pairwise error probability P_n we will use a result first presented in [Box54] and developed further in [Imh61] and [MP92, Section 4.3b]. Therein the PDF of central, indefinite quadratic forms in real Gaussian variables is derived. In order to use this result, we first need to do some preprocessing to get the model in a form that matches that of the cited references. To this end, we first express the quadratic form in complex variables in (10.61) as a quadratic form in real variables. Assume first that $|\mathbf{R}_0| < |\mathbf{R}_n|$ and define

$$\mathbf{z} = \begin{bmatrix} \mathrm{Re}\{\mathbf{y}\} \\ \mathrm{Im}\{\mathbf{y}\} \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} \mathrm{Re}\{\mathbf{\Psi}\} & -\mathrm{Im}\{\mathbf{\Psi}\} \\ \mathrm{Im}\{\mathbf{\Psi}\} & \mathrm{Re}\{\mathbf{\Psi}\} \end{bmatrix}. \tag{10A.1}$$

The quadratic form in (10.61) can then be expressed as

$$q \triangleq \mathbf{y}^H \mathbf{\Psi} \mathbf{y} = \mathbf{z}^T \mathbf{T} \mathbf{z}, \tag{10A.2}$$

where, now, the random vector $\mathbf{z}$ is real. The moments of $\mathbf{z}$ are

$$\mathrm{E}[\mathbf{z}] = \mathbf{0}, \tag{10A.3}$$

$$\mathrm{E}\left[\mathbf{z}\mathbf{z}^T\right] = \mathbf{\Sigma} = \frac{1}{2} \begin{bmatrix} \mathrm{Re}\{\mathbf{I}_N \otimes \mathbf{R}_0\} & -\mathrm{Im}\{\mathbf{I}_N \otimes \mathbf{R}_0\} \\ \mathrm{Im}\{\mathbf{I}_N \otimes \mathbf{R}_0\} & \mathrm{Re}\{\mathbf{I}_N \otimes \mathbf{R}_0\} \end{bmatrix}. \tag{10A.4}$$

Now, let the eigendecomposition of $\mathbf{\Sigma}^{1/2}\mathbf{T}\mathbf{\Sigma}^{1/2}$ be

$$\mathbf{\Sigma}^{1/2}\mathbf{T}\mathbf{\Sigma}^{1/2} = \mathbf{U}\mathbf{\Gamma}\mathbf{U}^T. \tag{10A.5}$$

The transformation

$$\mathbf{w} = \mathbf{U}^T \mathbf{\Sigma}^{-1/2} \mathbf{z} \tag{10A.6}$$

then brings the quadratic form into the canonical form

$$q = \sum_{j=1}^{2KN} \tilde{\gamma}_j w_j^2, \tag{10A.7}$$

where $\tilde{\gamma}_j, j = 1, \ldots, 2KN$, are the eigenvalues of $\mathbf{\Sigma}^{1/2}\mathbf{T}\mathbf{\Sigma}^{1/2}$ and w_j is the corresponding element in $\mathbf{w}$. The moments of $\mathbf{w}$ are

$$\mathrm{E}[\mathbf{w}] = \mathbf{0} \tag{10A.8}$$

$$\mathrm{E}\left[\mathbf{w}\mathbf{w}^T\right] = \mathbf{I}. \tag{10A.9}$$

Thus, the quadratic form has been written as a linear combination of independent χ^2 variables. Due to the structure of the quadratic form, some of the eigenvalues will be zero and some will be equal. The quadratic form can then be written as a linear combination of a reduced number of independent χ^2 variables with degrees of freedom given by the multiplicity of the corresponding eigenvalue, i.e.

$$q \sim \sum_{j=1}^{m} \gamma_j \chi^2_{\alpha_j}. \tag{10A.10}$$

Here, $\gamma_j, j = 1, \ldots, m$, are the distinct non-zero eigenvalues, α_j is the multiplicity of the j-th eigenvalue. Furthermore, χ^2_n denotes a χ^2 random variable with n degrees of freedom, and $\sim$ means that the left and right hand sides have the same distribution. The result in the above cited references is based on the assumption that all multiplicities of the eigenvalues are even numbers. To show that this is the case for the eigendecomposition in (10A.5) we present the following result, which relates the eigenvalues γ_j to the eigenvalues of $\mathbf{R}_0\mathbf{R}_n^{-1}$.

Lemma 10.1 *Let $d_i, i = 1, \ldots, m$, be the distinct eigenvalues of $\mathbf{R}_0\mathbf{R}_n^{-1}$ that satisfy $d_i \neq 1$, and denote the corresponding multiplicities by n_i. Assume that $|\mathbf{R}_0| < |\mathbf{R}_n|$. Then, the m distinct, non-zero eigenvalues of $\mathbf{\Sigma}^{1/2}\mathbf{T}\mathbf{\Sigma}^{1/2}$ are given by $\gamma_i = (d_i - 1)/2$ and they occur with multiplicities $2Nn_i$.*

Proof: First, we show the relationship between the eigenvalues of the complex Hermitian matrix $\mathbf{\Psi}$ and the real symmetric matrix $\mathbf{T}$. Let ξ be an eigenvalue and $\mathbf{e}$ be the corresponding eigenvector of $\mathbf{\Psi}$. Then, by definition,

$$\mathbf{\Psi e} = \xi \mathbf{e} \tag{10A.11}$$

so that

$$\mathrm{Re}\{\mathbf{\Psi e}\} = \xi \mathrm{Re}\{\mathbf{e}\} \tag{10A.12}$$
$$\mathrm{Im}\{\mathbf{\Psi e}\} = \xi \mathrm{Im}\{\mathbf{e}\}, \tag{10A.13}$$

since the eigenvalues of $\mathbf{\Psi}$ are real ($\mathbf{\Psi}$ being Hermitian). We will now show that

$$\mathbf{v}_1 = \begin{bmatrix} \mathrm{Re}\{\mathbf{e}\} \\ \mathrm{Im}\{\mathbf{e}\} \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} \mathrm{Im}\{\mathbf{e}\} \\ -\mathrm{Re}\{\mathbf{e}\} \end{bmatrix} \tag{10A.14}$$

are both eigenvectors of $\mathbf{T}$ with the same eigenvalue ξ. Using the definition of $\mathbf{T}$ in (10A.1) we obtain

$$\begin{aligned}\mathbf{T}\mathbf{v}_1 &= \begin{bmatrix} \mathrm{Re}\{\mathbf{\Psi}\} & -\mathrm{Im}\{\mathbf{\Psi}\} \\ \mathrm{Im}\{\mathbf{\Psi}\} & \mathrm{Re}\{\mathbf{\Psi}\} \end{bmatrix} \begin{bmatrix} \mathrm{Re}\{\mathbf{e}\} \\ \mathrm{Im}\{\mathbf{e}\} \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{Re}\{\mathbf{\Psi}\}\mathrm{Re}\{\mathbf{e}\} - \mathrm{Im}\{\mathbf{\Psi}\}\mathrm{Im}\{\mathbf{e}\} \\ \mathrm{Im}\{\mathbf{\Psi}\}\mathrm{Re}\{\mathbf{e}\} + \mathrm{Re}\{\mathbf{\Psi}\}\mathrm{Im}\{\mathbf{e}\} \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{Re}\{\mathbf{\Psi}\mathbf{e}\} \\ \mathrm{Im}\{\mathbf{\Psi}\mathbf{e}\} \end{bmatrix} = \xi \begin{bmatrix} \mathrm{Re}\{\mathbf{e}\} \\ \mathrm{Im}\{\mathbf{e}\} \end{bmatrix} = \xi\mathbf{v}_1. \end{aligned} \tag{10A.15}$$

A similar calculation shows that $\mathbf{T}\mathbf{v}_2 = \xi\mathbf{v}_2$. Hence, $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors of $\mathbf{T}$ with eigenvalue ξ. It then follows that $\mathbf{T}$ has the same eigenvalues as $\mathbf{\Psi}$, but with double multiplicity.

Next, we show the relationship between the eigenvalues of $\mathbf{\Sigma}^{1/2}\mathbf{T}\mathbf{\Sigma}^{1/2}$ and those of $\mathbf{R}_0\mathbf{R}_n^{-1}$. It is easy to show by a similarity transformation that $\mathbf{\Sigma}^{1/2}\mathbf{T}\mathbf{\Sigma}^{1/2}$ and $\mathbf{\Sigma}\mathbf{T}$ have the same eigenvalues. From (10A.1) and (10A.3) we obtain

$$\begin{aligned}\mathbf{\Sigma}\mathbf{T} &= \frac{1}{2}\begin{bmatrix} \mathrm{Re}\{\mathbf{I}_N \otimes \mathbf{R}_0\} & -\mathrm{Im}\{\mathbf{I}_N \otimes \mathbf{R}_0\} \\ \mathrm{Im}\{\mathbf{I}_N \otimes \mathbf{R}_0\} & \mathrm{Re}\{\mathbf{I}_N \otimes \mathbf{R}_0\} \end{bmatrix} \begin{bmatrix} \mathrm{Re}\{\mathbf{\Psi}\} & -\mathrm{Im}\{\mathbf{\Psi}\} \\ \mathrm{Im}\{\mathbf{\Psi}\} & \mathrm{Re}\{\mathbf{\Psi}\} \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} \mathrm{Re}\{(\mathbf{I}_N \otimes \mathbf{R}_0)\mathbf{\Psi}\} & -\mathrm{Im}\{(\mathbf{I}_N \otimes \mathbf{R}_0)\mathbf{\Psi}\} \\ \mathrm{Im}\{(\mathbf{I}_N \otimes \mathbf{R}_0)\mathbf{\Psi}\} & \mathrm{Re}\{(\mathbf{I}_N \otimes \mathbf{R}_0)\mathbf{\Psi}\} \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} \mathrm{Re}\{\mathbf{I}_N \otimes (\mathbf{R}_0\mathbf{R}_n^{-1} - \mathbf{I}_K)\} & -\mathrm{Im}\{\mathbf{I}_N \otimes (\mathbf{R}_0\mathbf{R}_n^{-1} - \mathbf{I}_K)\} \\ \mathrm{Im}\{\mathbf{I}_N \otimes (\mathbf{R}_0\mathbf{R}_n^{-1} - \mathbf{I}_K)\} & \mathrm{Re}\{\mathbf{I}_N \otimes (\mathbf{R}_0\mathbf{R}_n^{-1} - \mathbf{I}_K)\} \end{bmatrix}. \end{aligned} \tag{10A.16}$$

The structure of $\mathbf{\Sigma}\mathbf{T}$ is therefore the same as that of $\mathbf{T}$ (regarding real and imaginary parts). From the previous calculations we can therefore conclude that $\mathbf{\Sigma}\mathbf{T}$ has the same eigenvalues as $\mathbf{I}_N \otimes (\mathbf{R}_0\mathbf{R}_n^{-1} - \mathbf{I}_K)/2$, but occurring with double multiplicity. The eigenvalues of $\mathbf{I}_N \otimes (\mathbf{R}_0\mathbf{R}_n^{-1} - \mathbf{I}_K)/2$ are those of $(\mathbf{R}_0\mathbf{R}_n^{-1} - \mathbf{I}_K)/2$ occurring with multiplicity Nn_i. Finally, the eigenvalues of $\mathbf{R}_0\mathbf{R}_n^{-1} - \mathbf{I}_K$ are $d_j - 1$, where d_j are the eigenvalues of $\mathbf{R}_0\mathbf{R}_n^{-1}$. The assertion in the lemma follows. □

We are now in the position to apply the result in [MP92, Section 4.3b]. Let the m distinct non-zero eigenvalues of $\mathbf{\Sigma}^{1/2}\mathbf{T}\mathbf{\Sigma}^{1/2}$ be ordered such that $\gamma_1 > \ldots > \gamma_p > 0 > \gamma_{p+1} > \ldots > \gamma_m$. The quadratic form can then be written as

$$q \sim \lambda_1\chi^2_{2Nn_1} + \ldots + \lambda_p\chi^2_{2Nn_p} - \lambda_{p+1}\chi^2_{2Nn_{p+1}} - \ldots - \lambda_m\chi^2_{2Nn_m} \tag{10A.17}$$

where

$$\lambda_i \;=\; \begin{cases} \gamma_i, & i = 1, \ldots, p \\ -\gamma_i, & i = p+1, \ldots, m \end{cases} \tag{10A.18}$$

and $2Nn_i$ is the multiplicity of γ_i. From [MP92, Section 4.3b] the PDF of the quadratic form q in (10A.17) is now obtained as

$$f(q) \;=\; \sum_{i=p+1}^{m} c_i e^{q/2\lambda_i} \sum_{\beta=0}^{Nn_i-1} \binom{Nn_i - 1}{\beta} q^{Nn_i-1-\beta} B_i^{(\beta)}, \quad q < 0, \tag{10A.19}$$

where c_i and $B_i^{(\beta)}$ are defined in (10.63) and (10.64), respectively. The PDF for $q \geq 0$ is also given in [MP92, Section 4.3b], but we do not need it here since we have arranged the calculations so that we need only integrate the PDF over negative q. The CDF of the quadratic form is now obtained by integrating the PDF. The result is

$$\begin{aligned} F(x) \;&=\; \Pr[q \leq x] = \int_{-\infty}^{x} f(q)dq \\ &=\; \sum_{i=p+1}^{m} c_i \sum_{\beta=0}^{Nn_i-1} \binom{Nn_i - 1}{\beta} B_i^{(\beta)} \int_{-\infty}^{x} q^{Nn_i-1-\beta} e^{q/2\lambda_i}\, dq, \quad x < 0. \end{aligned} \tag{10A.20}$$

Using [GR94]

$$\int_u^{\infty} t^n e^{-\mu t} dt \;=\; e^{-u\mu} \sum_{k=0}^{n} \frac{n!}{k!} \frac{u^k}{\mu^{n-k+1}}, \quad u > 0, \mathrm{Re}\{\mu\} > 0, \tag{10A.21}$$

we obtain (10.67). From (10.61), (10A.2), and (10A.20) we have

$$P_n \;=\; \Pr[q < N \log |\mathbf{R}_0 \mathbf{R}_n^{-1}|] = F(N \log |\mathbf{R}_0 \mathbf{R}_n^{-1}|). \tag{10A.22}$$

We need to check that $\log |\mathbf{R}_0 \mathbf{R}_n^{-1}| < 0$ since the CDF in (10A.20) is valid only for negative arguments. Since, by assumption, $|\mathbf{R}_0| < |\mathbf{R}_n|$ it follows immediately that $\log |\mathbf{R}_0 \mathbf{R}_n^{-1}| = \log(|\mathbf{R}_0|/|\mathbf{R}_n|) < 0$.

For the case $|\mathbf{R}_n| < |\mathbf{R}_0|$ we write the pairwise error probability as

$$\begin{aligned} P_n \;&=\; \Pr\left[\mathrm{Tr}\left\{\left[\mathbf{R}_n^{-1} - \mathbf{R}_0^{-1}\right] \hat{\mathbf{R}}\right\} < \log\left|\mathbf{R}_0 \mathbf{R}_n^{-1}\right|\right] \\ &=\; 1 - \Pr\left[\mathrm{Tr}\left\{\left[\mathbf{R}_0^{-1} - \mathbf{R}_n^{-1}\right] \hat{\mathbf{R}}\right\} < -\log\left|\mathbf{R}_0 \mathbf{R}_n^{-1}\right|\right] \\ &=\; 1 - F\left(-N \log\left|\mathbf{R}_0 \mathbf{R}_n^{-1}\right|\right), \end{aligned} \tag{10A.23}$$

where, now, $-\log |\mathbf{R}_0 \mathbf{R}_n^{-1}| < 0$. Otherwise, the only thing that differs in the proof of the $|\mathbf{R}_n| < |\mathbf{R}_0|$ case is that the eigenvalues of $\mathbf{\Sigma}^{1/2} \mathbf{T} \mathbf{\Sigma}^{1/2}$ will be $(1 - d_i)/2$. This completes the proof of Theorem 10.2. □

Chapter 11

Application Examples

In this chapter we present two application examples which are concerned with DOA estimation using sparse arrays. The inherently high sidelobes of such arrays make them prone to ambiguities in the estimation. Therefore, analyzing the threshold effect is particularly important for this class of arrays. The examples are edited versions of previously published conference papers, [AE00, Ath01b], and are presented here in two self-contained sections. Since these papers were published before we developed the general analysis presented in the preceding chapters, the focus in this chapter is on lower bounds. However, in the first example we also present an analysis that is similar to the previous one, but is only applicable to a particular array structure.

11.1 DOA Estimation Using Separated Subarrays

The first example is concerned with DOA estimation with a particular class of sparse linear arrays, characterized by two widely separated subarrays. Since a large array aperture is obtained with a small number of elements, this structure can provide very accurate angle estimates at a reasonable cost, but at the expense of near-ambiguities. The fundamental limitations in DOA estimation with this class of arrays are investigated. We evaluate the performance of DOA estimation algorithms and compare with theoretical bounds, taking the effects of near-ambiguities into account.

11.1.1 Introduction

Sensor arrays are used for DOA estimation in a wide range of applications such as radar, passive sensors for jammer localization, sonar, wireless communications, seismic analysis and medical imaging. The perhaps most important parameter when designing sensor arrays is the array size. Large arrays with many sensors can provide very accurate DOA estimates. The cost of such a system, however, grows with the number of sensors, both in terms of receiver hardware and computational complexity. A way to reduce the cost while maintaining accurate DOA estimates is to reduce the number of sensors without decreasing the array size. In such sparse arrays, the distance between adjacent sensors will exceed half a wavelength of the impinging wavefronts. This will lead to grating lobes, or near-ambiguities, in the array beampattern. These grating lobes may produce large estimation errors if potential ambiguities are not resolved. In order for sparse arrays to be useful for DOA estimation, the probability of unresolved ambiguities must be very low. This can be achieved by careful array design (see, e.g., [VE99]) or by using estimation algorithms that are specifically designed to resolve the potential ambiguities. DOA estimation algorithms designed to resolve ambiguities are often based on a combination of two estimates, one coarse non-ambiguous and one fine ambiguous estimate [Bag89, JR81, LvdVD99, WZ98, ZW00]. The coarse estimate is used to disambiguate the fine estimate to obtain a fine non-ambiguous estimate. Other ways of achieving disambiguation include frequency diversity [Bom00] and multiple target tracking [WT00].

In this example we focus on a particular class of sparse linear arrays. The array is an interferometric-like system composed of two spatially separated subarrays. The distance between the phase centers of the subarrays is D. There is a spatial gap between the subarrays that is very large compared to the usual inter-sensor spacing of half a wavelength, see Figure 11.1. This will lead to a large number of grating lobes in the array beampattern with magnitudes close to the magnitude of the mainbeam. Each subarray is a ULA, with a distance $d \leq \lambda/2$ between adjacent sensors, where λ is the wavelength of the impinging wavefront. One might question the choice of such an array structure. As will be shown later, it is nearly optimal from an estimation point of view when the SNR is high. At low SNR, this geometry is a poor choice, since the high sidelobes will give large errors. However, incorporating implementation and manufacturing aspects, this structure appears attractive for many applications. The regular structure of the subarrays means that they can be mass produced at low cost. These subarrays can then for instance be placed far apart on the hull of a vessel to provide a large baseline for very accurate DOA estimation. Of course, one could use more than two

separated subarrays, which would be good for the disambiguation. However, for simplicity and ease of exposition only the case of two subarrays is considered herein.

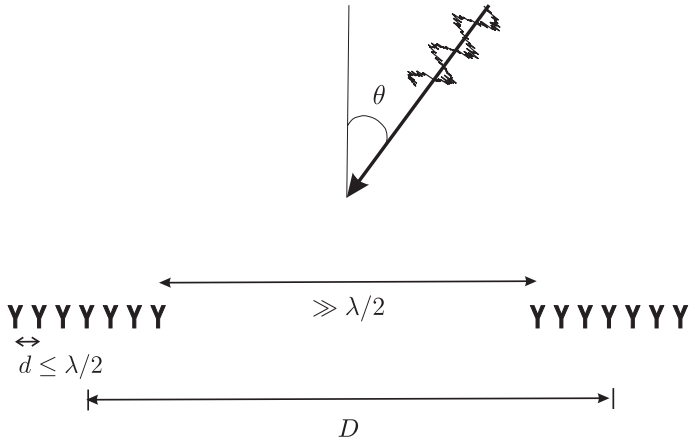

Figure 11.1: Array configuration

Figure 11.2 shows the array beampattern for two subarrays separated a distance $33\lambda/2$ between the phase centers of the subarrays. There are 4 sensors in each subarray. The shape of the beampattern illustrates the

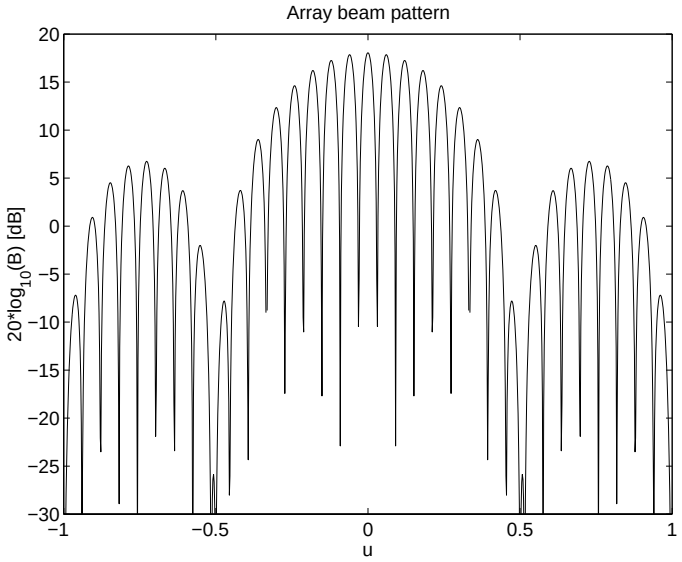

Figure 11.2: Array beampattern in dB

difficulty in using this sensor arrangement for DOA estimation. The narrow

mainlobe indicates that a high accuracy can be achieved, but the many and high sidelobes indicate a great risk for large errors.

The array beampattern is the product of the array factor and the subarray beampattern. The array factor is the beampattern that would result from two isotropic sensors placed at the phase centers of the subarrays. This pattern has grating lobes of identical heights as the mainlobe. The subarrays have the effect of reducing the grating lobes in the overall beampattern by multiplying the array factor with the subarray beampattern. In Chapter 1, a grating lobe was defined as a lobe of the same height as the mainlobe that appears due to spatial aliasing. In this sense, the overall beampattern in Figure 11.2 has no grating lobes, since the sidelobe peaks are slightly lower than the main peak due to multiplication by the subarray pattern. Still, we call these lobes grating lobes since the reason for their appearance is spatial aliasing. If the grating lobes were of exactly the same height as the mainlobe, it would be impossible to resolve the ambiguities without a priori information about the DOA. One way to interpret DOA estimation with an array that has such a structure is that the long baseline between the subarrays gives a very accurate but ambiguous DOA estimate. The individual subarrays can then be used to determine which of the many peaks in the large baseline beampattern that is the correct one.

Relation to Previous Work

Several authors have treated various, more narrow, aspects of the present problem, either treating performance bounds, or presenting specific algorithms for array structures similar to the one considered here. In [ZS98] the Cramér-Rao Bound (CRB) was compared with the Weiss-Weinstein Bound (WWB), in [LvdVD99, WZ98, ZW00] estimation algorithms specifically designed to resolve ambiguities were presented, and in [BS01] the Maximum Likelihood (ML) estimator and the Barankin bound were computed for separated subarrays. A recent study on two or more separated subarrays has been reported in [Zat03].

Aim of Present Work

The aim of the work presented herein is to explore further the fundamental limitations in ambiguity prone DOA estimation with this particular array structure. To this end, optimal DOA estimation is developed and compared with performance bounds. A simple approximate expression for the ultimate estimation accuracy that can be achieved with this class of arrays is also derived.

11.1.2 Data Model and Problem Formulation

We consider the problem of finding the DOA of a single signal impinging on an array structure as depicted in Figure 11.1. For mathematical convenience we consider estimation of $u \triangleq \sin\theta$, where θ is defined to be the angle w.r.t. boresight[1] of the array. The array output at time t is modeled by the $2K \times 1$ complex vector

$$\mathbf{x}(t) = \mathbf{a}(u)s(t) + \mathbf{n}(t), \qquad t = 1, \ldots, N \tag{11.1}$$

where $\mathbf{a}(u)$ is a $2K \times 1$ steering vector that models the array response to a unit waveform from direction u, K is the number of sensors in one subarray, N is the number of snapshots, $s(t)$ is the complex amplitude at time t at baseband of the impinging wavefront and $\mathbf{n}(t)$ is the noise. The signal $s(t)$ and noise $\mathbf{n}(t)$ are assumed independent and they are modeled as white (spatially and temporally), zero mean, circular complex Gaussian random variables with second order moments

$$\begin{aligned} \mathrm{E}\left[|s(t)|^2\right] &= S & (11.2)\\ \mathrm{E}[\mathbf{n}(t)\mathbf{n}^H(s)] &= \mathbf{I}\,\delta_{t,s} & (11.3)\\ \mathrm{E}[\mathbf{n}(t)\mathbf{n}^T(s)] &= \mathbf{0} \quad \forall\, t, s & (11.4) \end{aligned}$$

where S is assumed to be known. The signal variance S thus corresponds to the Signal-to-Noise-Ratio (SNR) per space-time sample since the noise variance is normalized to unity. Although the assumption that S is known may be unrealistic, we will later find that this is not particularly restrictive in this example. The two subarrays are assumed to be ideal and identical with omnidirectional antenna elements. The subarray steering vectors are then given by

$$\mathbf{a}_1(u) = \mathbf{a}_2(u) = \begin{bmatrix} 1 & e^{-j2\pi d_n u} & \ldots & e^{-j(K-1)2\pi d_n u} \end{bmatrix}^T, \tag{11.5}$$

where $d_n \triangleq d/\lambda$ is the inter-sensor spacing in units of wavelengths. The steering vector for the whole array is then given by

$$\mathbf{a}(u) = \begin{bmatrix} \mathbf{a}_1(u) \\ e^{-j2\pi D_n u}\,\mathbf{a}_1(u) \end{bmatrix}, \tag{11.6}$$

where $D_n \triangleq D/\lambda$. With these assumptions we obtain the following model for the data covariance matrix

$$\mathbf{R} = \mathrm{E}\left[\mathbf{x}(t)\mathbf{x}^H(t)\right] = S\mathbf{a}(u)\mathbf{a}^H(u) + \mathbf{I}. \tag{11.7}$$

[1] Boresight means perpendicular to the array axis.

The problem considered in this example is to estimate $u = \sin\theta$ based on the noisy observations $\mathbf{X} \triangleq \begin{bmatrix} \mathbf{x}(1) & \dots & \mathbf{x}(N) \end{bmatrix}$. Particular attention is paid to the problem of near ambiguities due to the grating lobes in the array beampattern. First, we seek lower bounds on the estimation error variance that take the effects of near ambiguities into account.

11.1.3 Performance Bounds

This section is concerned with lower bounds on estimation accuracy. It is shown that the CRB is not suitable for predicting the attainable performance of DOA estimation with separated subarrays. Indeed, the CRB would suggest that the variance could be made arbitrarily small just by using a sufficiently large separation. Therefore, other bounds that are available in the literature are applied to the present problem. Both classical and Bayesian bounds are treated, and it is argued that they can be compared with one another under certain circumstances.

Classical and Bayesian Mean Square Errors

In the following, both classical and Bayesian bounds will be treated. Normally, it makes no sense to compare classical and Bayesian bounds, since the underlying experiments are inherently different. For the problem at hand, however, some circumstances make such a comparison meaningful. These circumstances will be explained in more detail later. We seek lower bounds on the classical MSE

$$\mathrm{CMSE}(\hat{u}, u) = \mathrm{E}_{\mathbf{X}}\left[(\hat{u}-u)^2\right] = \int (\hat{u}-u)^2 \, p(\mathbf{X}; u) d\mathbf{X} \tag{11.8}$$

and the Bayesian MSE

$$\begin{aligned} \mathrm{BMSE}(\hat{u}) &= \mathrm{E}_{\mathbf{X},u}\left[(\hat{u}-u)^2\right] = \iint (\hat{u}-u)^2 \, p(\mathbf{X}, u) d\mathbf{X} du \\ &= \iint (\hat{u}-u)^2 \, p(\mathbf{X}|u) p(u) d\mathbf{X} du \end{aligned} \tag{11.9}$$

where E denotes expectation and $\hat{u}$ is an estimate of u based on the observations $\mathbf{X} = \begin{bmatrix} \mathbf{x}(1) & \dots & \mathbf{x}(N) \end{bmatrix}$. The likelihood function $p(\mathbf{X}; u)$ is the PDF of $\mathbf{X}$ as a function of the *deterministic* parameter u and $p(\mathbf{X}, u)$ is the joint PDF of $\mathbf{X}$ and the *random* parameter u. The density $p(\mathbf{X}|u)$ is the PDF of $\mathbf{X}$ conditioned on u and $p(u)$ is the prior PDF of the random parameter u. Although $p(\mathbf{X}; u)$ and $p(\mathbf{X}|u)$ have different interpretations, they will have identical mathematical forms, a fact that we will make use of later on. By

$d\mathbf{X}$ we mean $d\mathbf{x}(1)\ d\mathbf{x}(2)\cdots d\mathbf{x}(N)$. We will also refer to the classical MSE as the *local* MSE, since it represents the MSE for a particular value of u. Accordingly, the Bayesian MSE will be referred to as the *global* MSE since it is an average of the local MSE over the prior PDF.

Obviously, some constraint must be imposed on the estimator when the bound on the CMSE is considered. Otherwise, the trivial bound of zero would be obtained if we simply let $\hat{u} = u$. Usually, bounds on the CMSE are restricted to the class of unbiased estimators. For the DOA estimation problem, an unbiased estimator cannot exist since the parameter support is finite [ZZ69]. However, we have found in simulations that for the cases of interest in this example, the squared bias is very small compared to the variance. Therefore, we will not discuss the issue of bias any further herein.

The A Priori Distribution

Throughout this chapter, when treating u as random parameter, a uniform prior PDF on $[-u_0, u_0]$ will be assigned to u. This corresponds to a cosine distribution of the DOA

$$p(\theta) = \frac{1}{2u_0}\cos\theta \qquad |\theta| \leq \arcsin u_0 \tag{11.10}$$

which means that DOAs close to boresight are more likely than DOAs near endfire[2]. Although the reason for assigning a uniform prior PDF to u was mathematical convenience rather than physical considerations, some justification can be given. One could e.g. argue that in some applications the array boresight is mechanically aimed at a certain direction where targets are most likely to appear in a given scenario. Another motivation could be that in a practical sensor array the individual sensors usually have a directional beampattern. Although this variation in array gain has not been accounted for in the data model, it can provide justification for the assumed prior distribution.

Handling Edge Effects

If the inter-sensor spacing is $\lambda/2$ and $u_0 = 1$, edge effects may cause troubles in evaluation of the Bayesian MSE. If u is slightly less than +1, it may be estimated as a value slightly greater than -1 due to the cyclical nature of the beampattern. This may cause the estimated MSE to become "unfairly" large. One way to get around this problem is to "unwrap" the estimation error by calculating it modulo-2. This re-definition of the estimation error requires

[2]Endfire means parallel to the array axis.

that the bounds are modified accordingly, see [BB00] for a modification of the Ziv-Zakai bound to this case. We will not take this approach herein. Instead, the problem is avoided by ascertaining that these types of errors never occur. The problem is the nature of $\exp(-j2\pi d_n u)$ in the array steering vector when $d_n u$ is close to 0.5. Clearly, it is difficult to tell the difference between $\exp[j\pi(1-\epsilon)]$ and $\exp[-j\pi(1-\epsilon)]$, where ϵ is a small number, from noisy observations. If the phase error is measured modulo 2π, it will be very small, if not it will be close to 2π. By assuring that $d_n u$ never comes close to 0.5, we can avoid this problem. This can be achieved either by letting $d_n < 0.5$, which corresponds to spatial oversampling, or by letting $u_0 < 1$, i.e. constraining the parameter range by the prior PDF.

General Optimal Estimation

In the classical case, there is no general procedure for finding the minimum variance unbiased estimator. In the Bayesian case however, it is well known [Tre68] that the Minimum MSE (MMSE) estimator is given by the mean of the posterior PDF

$$\hat{u} = \mathrm{E}[u|\mathbf{X}] = \int up(u|\mathbf{X})du. \tag{11.11}$$

The MSE of this estimator is thus a tight lower bound. Since evaluation of this MSE is computationally intractable, a number of bounds which are easier to compute have been reported in the literature. In this chapter we will study the Weiss-Weinstein bound (WWB) which we will compare with other bounds and results from Monte Carlo simulations. The subsequent sections will provide expressions for the Cramér-Rao, Barankin and Weiss-Weinstein bounds for the problem at hand. The chapter ends with a comparison of these different bounds.

The Cramér-Rao Bound

Two different CRBs will be used to illustrate different modes of operation. Above the ambiguity threshold, the data from the subarrays can be combined (spatially) coherently to obtain DOA estimates based on the full aperture. We call the CRB for this case the coherent CRB and denote it B_C. Below the ambiguity threshold, the data from the subarrays can be combined only incoherently. In this case we assume that we have available two independent DOA estimates, one from each subarray, making use only of the subarray aperture. We call the CRB for this case the incoherent CRB and denote it B_I. In Appendix 11A it is shown that the coherent and incoherent CRBs are

given by

$$\begin{aligned} B_C &= \frac{3(2SK+1)}{8\pi^2 NS^2K^2\left(d_n^2\left(K^2-1\right)+3D_n^2\right)} \approx \frac{3}{4\pi^2 NKS\left(d_n^2K^2+3D_n^2\right)} \\ B_I &= \frac{3(SK+1)}{4\pi^2 d_n^2 NS^2K^2\left(K^2-1\right)} \approx \frac{3}{4\pi^2 d_n^2 NSK^3}. \end{aligned} \tag{11.12}$$

The coherent CRB is thus a factor $1+3D_n^2/d_n^2K^2$ lower than the incoherent CRB. If $3D_n^2/d_n^2K^2 \gg 1$ this factor is $3D^2/L^2$ where $L=Kd$ is the effective length of a subarray. This means that if the SNR is sufficient for neglecting ambiguity errors, the standard deviation of the estimation error using separated subarrays is reduced by a factor which is proportional to the distance between the subarray phase centers divided by the effective length of one subarray.

The Barankin Bound

The Barankin bound (BB) for an underwater matched-field processing problem with a model similar to the one considered here has been derived in [TK99]. We use their result adapted to our model. The BB for the classical MSE of any unbiased estimator of u from the data vector $\mathbf{x}$ is given by

$$\text{CMSE} \geq \text{BB} = \mathbf{h}\left(\mathbf{B}-\mathbf{1}\mathbf{1}^T\right)^{-1}\mathbf{h}^T \tag{11.13}$$

where $\mathbf{1}$ is a column vector of ones and $\mathbf{h}$ is the $1\times T$ row vector

$$\mathbf{h} = \left[\begin{array}{ccc} u_1-u & \cdots & u_T-u \end{array}\right], \tag{11.14}$$

where $u_i,\ i=1,\ldots,T$ are referred to as test points. They represent perturbations in the true parameter value u. If the test points are selected judiciously, the number of test points T can be relatively few. The test points should be chosen so as to maximize the right hand side of (11.13) in order to obtain as tight a bound as possible. A common choice is to pick the test points as the parameter values that correspond to ambiguity peaks in the criterion function, see e.g. [MS69]. We therefore choose the test points as the values of u corresponding to the mainlobe and the nearest grating lobes. The original formulation of the BB is actually more general. For details concerning this general form, see [Bar49]. The elements of the $\mathbf{B}$ matrix is given by

$$\mathbf{B}_{ij} = \mathrm{E}\left[L(\mathbf{X},u_i,u)L(\mathbf{X},u_j,u)\right] \quad i,j=1,\ldots,T, \tag{11.15}$$

where $L(\mathbf{X},u_i,u)$ is the likelihood ratio

$$L(\mathbf{X},u_i,u) = \frac{p(\mathbf{X}|u_i)}{p(\mathbf{X}|u)}. \tag{11.16}$$

The elements in the $\mathbf{B}$ matrix can be shown to be given by [TK99]

$$\mathbf{B}_{ij} = \left(\frac{|\mathbf{R}(u)|}{|\mathbf{R}(u_i)| \cdot |\mathbf{R}(u_j)| \cdot |\mathbf{R}^{-1}(u_i) + \mathbf{R}^{-1}(u_j) - \mathbf{R}^{-1}(u)|} \right)^N, \quad (11.17)$$

which can be simplified to

$$\begin{aligned} \mathbf{B}_{ij} &= \left[\frac{1+4SK}{(1+2SK)^2} + 2(1+2SK)\mathrm{Re}\{\tilde{\mathbf{a}}^H\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^H\tilde{\mathbf{a}}_j\tilde{\mathbf{a}}_j^H\tilde{\mathbf{a}}\} \right. \\ &+ \left. |\tilde{\mathbf{a}}^H\tilde{\mathbf{a}}_i|^2 + |\tilde{\mathbf{a}}^H\tilde{\mathbf{a}}_j|^2 - (1+4SK)\,|\tilde{\mathbf{a}}_i^H\tilde{\mathbf{a}}_j|^2 \right]^{-N}, \end{aligned} \quad (11.18)$$

where

$$\tilde{\mathbf{a}}_i = \sqrt{\frac{S}{1+2SK}}\mathbf{a}(u_i). \quad (11.19)$$

The Barankin bound can then be evaluated by substituting (11.18) into (11.13).

The Weiss-Weinstein Bound

The WWB states that [WW85]

$$\mathrm{BMSE} \geq \mathrm{WWB} = \mathbf{h}\boldsymbol{\Gamma}^{-1}\mathbf{h}^T, \quad (11.20)$$

where $\mathbf{h}$ is the $1 \times T$ row vector of test points

$$\mathbf{h} = \begin{bmatrix} h_1 & \cdots & h_T \end{bmatrix} \quad (11.21)$$

and

$$\boldsymbol{\Gamma} = E[\mathbf{w}\mathbf{w}^T], \quad (11.22)$$

where

$$w_i \triangleq \frac{L^{1/2}(\mathbf{X}; u+h_i, u) - L^{1/2}(\mathbf{X}; u-h_i, u)}{\mathrm{E}[L^{1/2}(\mathbf{X}; u+h_i, u)]}, \quad (11.23)$$

is the i-th element in the $T \times 1$ real random vector $\mathbf{w}$. The likelihood ratio $L(\mathbf{X}; u+h, u)$ is given by

$$L(\mathbf{X}; u+h, u) \triangleq \frac{p(\mathbf{X}, u+h)}{p(\mathbf{X}, u)}. \quad (11.24)$$

The original formulation of the WWB is actually more general. For details concerning this general form, see [WW85]. Again, the test points are chosen from the grating lobe positions in the array beampattern.

In [DeL93], the WWB on azimuth and elevation estimation using a general planar array is derived. A slight modification of those calculations gives that the (i,j)-th element in $\mathbf{\Gamma}$ for the problem at hand is given by

$$\mathbf{\Gamma}_{ij} = 2\frac{C(h_i,h_j) - C(h_i,-h_j)}{C(h_i,0)C(h_j,0)}, \tag{11.25}$$

where

$$C(h_i,h_j) = A(h_i,h_j)B(h_i,h_j). \tag{11.26}$$

Here,

$$A(h_i,h_j) = \left[1 + \frac{(2KS)^2}{4(2KS+1)}\left(1 - \frac{\left|\mathbf{a}^H(h_i)\mathbf{a}(h_j)\right|^2}{4K^2}\right)\right]^{-N}, \tag{11.27}$$

$$B(h_i,h_j) = 1 - \frac{1}{2u_0}\max\left\{|h_i|,|h_j|,|h_i-h_j|\right\}. \tag{11.28}$$

The Bayesian Cramér-Rao Bound

The Bayesian CRB requires some regularity conditions which are not satisfied when a uniform prior PDF is assigned to u (the prior PDF is not twice differentiable). However, by smoothing the edges of the uniform distribution one can show [NV94] that the Bayesian CRB approaches the classical CRB at high SNR. We will therefore use the classical CRB when we compare with other bounds and with simulations.

Comparison of Bounds

In this section, the different bounds described in the previous sections will be compared. As noted in the introduction, comparing classical and Bayesian bounds is usually not meaningful. There are, however, some circumstances in the present study that make such a comparison meaningful. The key is that the local MSE is approximately independent of u. From (11.12) it is seen that the CRB is independent of u. Later on, we will find by simulation that also the optimal Bayesian estimator will have a local MSE which is approximately independent of u, when it is evaluated in the classical sense. By this we mean that the Bayesian philosophy is used to derive the optimal estimator. This estimator is then evaluated in a Monte Carlo simulation as

if u were deterministic, i.e. u is fixed from trial to trial. The local (classical) MSE is then computed for this particular value of u. This property, that the local MSE is independent of u, means that it does not matter if we evaluate the performance of an estimator in the classical or in the Bayesian sense. To see this, compare the local MSE

$$\mathrm{CMSE}(\hat{u}, u) = \int (\hat{u} - u)^2 \; p(\mathbf{X}; u) \, d\mathbf{X} \tag{11.29}$$

with the global (Bayesian) MSE

$$\mathrm{BMSE}(\hat{u}) = \iint (\hat{u} - u)^2 \, p(\mathbf{X}|u) \, p(u) \, d\mathbf{X} \, du = \int \mathrm{CMSE}(\hat{u}, u) \, p(u) \, du \tag{11.30}$$

since $p(\mathbf{X}; u)$ and $p(\mathbf{X}|u)$ have identical mathematical forms. Now, if $\mathrm{CMSE}(\hat{u}, u)$ is independent of u we obtain

$$\mathrm{BMSE}(\hat{u}) = \mathrm{CMSE}(\hat{u}) \int p(u) du = \mathrm{CMSE}(\hat{u}). \tag{11.31}$$

Although there might be some philosophical controversy in comparing classical and Bayesian estimators, the above calculations show that the local and global MSEs are the same if the local MSE is independent of the parameter value. Therefore, the MSE of the Bayesian MMSE estimator is also a lower bound on the MSE for the class of classical estimators that have an MSE which is independent of the parameter value. Suppose now that the Bayesian MMSE estimator, when evaluated in the classical sense, has a local MSE which is independent of the parameter value. If this is the case, this estimator is also the MMSE estimator in the classical sense for the class of estimators that have an MSE which is independent of the parameter value. In the next chapter, we will find by simulation that the Bayesian MMSE estimator has a local MSE that is approximately independent of u if the SNR is sufficiently high and if the edge effects discussed in a previous section are neglected.

In Figure 11.3, the BB, WWB, the incoherent and coherent CRBs are plotted versus the subarray separation D_n. For small subarray separation, the BB, WWB and coherent CRB coincide. In this region, ambiguity errors have no effect on the bounds. The MSE decreases as the subarray separation increases up to a certain point when ambiguous estimates begin to affect the bounds. We call this point the *ambiguity threshold*. Beyond this point the MSE starts to increase since ambiguous estimates occur more often. This effect is captured by the BB and WWB, but not by the CRB. At very large

subarray separation, the BB and WWB approach the incoherent CRB. In this region, only the envelope of the array beampattern provides useful information. It is interesting to note that there is indeed an optimal subarray separation that minimizes the MSE. A system that is able to operate just to the left of this threshold is the optimal design and provides a significant reduction of the MSE compared to a standard ULA with corresponding number of sensors. It can also be seen from the figure that the WWB is a much tighter bound than the BB. Therefore, the BB is not considered any further in this chapter. In the next section, the Bayesian MMSE estimator is implemented and its MSE is compared with the WWB in order to investigate the tightness of the WWB.

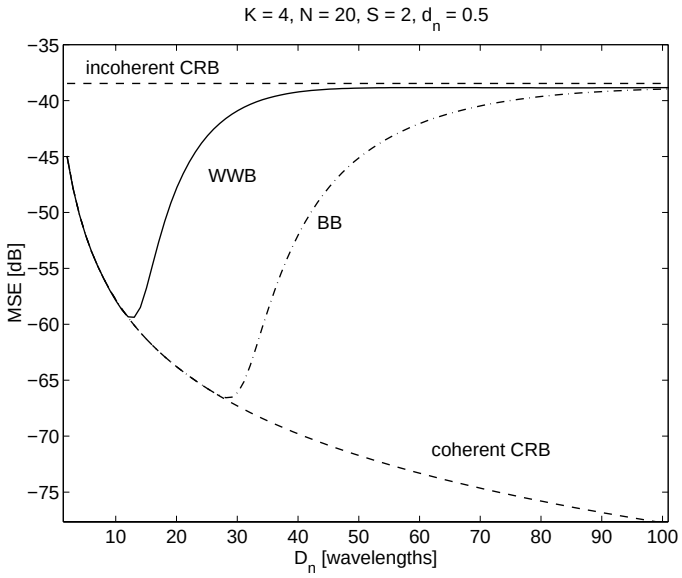

Figure 11.3: Bounds on MSE vs subarray separation

11.1.4 Performance of Estimators

In this chapter, the optimal Bayesian estimator for the problem at hand is derived. We also develop the Maximum A Posteriori (MAP) estimator and show its relation to the classical ML estimator and the conventional beamformer. A simple, approximate expression for the MSE of the MAP estimator is also derived. The performance of the different estimators are evaluated by means of Monte Carlo simulations. This is compared to the performance predicted by the approximation of the MAP MSE and the bounds treated in the previous section.

Optimal Bayesian Estimation

The optimal (in the Bayesian MSE sense) MMSE estimator is given by the mean of the posterior PDF (see e.g. [Tre68])

$$\hat{u} = \mathrm{E}[u|\mathbf{X}] = \int_U up(u|\mathbf{X})du \tag{11.32}$$

To compute the MMSE estimator, the posterior PDF of the parameter u given the data $\mathbf{X}$ is needed. From Bayes' rule we have

$$p(u|\mathbf{X}) = \frac{p(\mathbf{X}|u)p(u)}{p(\mathbf{X})} = \frac{p(\mathbf{X}|u)p(u)}{\displaystyle\int_U p(\mathbf{X}|u)p(u)du}. \tag{11.33}$$

We assume u has a uniform prior PDF on $[-u_0, u_0]$ so that $p(u) = 1/(2u_0)$ for $-u_0 \leq u \leq u_0$ and $p(u) = 0$ elsewhere. The posterior PDF then simplifies to

$$p(u|\mathbf{X}) = \frac{p(\mathbf{X}|u)}{\displaystyle\int_{-u_0}^{u_0} p(\mathbf{X}|u)du} \tag{11.34}$$

The MAP estimator is thus identical in form to the ML estimator when the prior PDF is uniform. It also follows that the MMSE estimator is given by

$$\hat{u}_{\mathrm{MMSE}} = \frac{\displaystyle\int_U up(\mathbf{X}|u)du}{\displaystyle\int_U p(\mathbf{X}|u)du} \tag{11.35}$$

which is recognized as the center of gravity of the conditional density $p(\mathbf{X}|u)$. This conditional PDF $p(\mathbf{X}|u)$ is in the white Gaussian case given by

$$\begin{aligned} p(\mathbf{X}|u) &= \prod_{t=1}^{N} \frac{1}{\pi^{2K}|\mathbf{R}|} \exp\left[-\mathbf{x}^H(t)\mathbf{R}^{-1}\mathbf{x}(t)\right] \\ &= \frac{1}{\pi^{2KN}|\mathbf{R}|^N} \exp\left[-\sum_{t=1}^{N} \mathbf{x}^H(t)\mathbf{R}^{-1}\mathbf{x}(t)\right] \end{aligned} \tag{11.36}$$

where $|\mathbf{R}|$ denotes the determinant of $\mathbf{R}$. Applying the matrix inversion lemma [Lüt96] to (11.7) gives

$$\mathbf{R}^{-1} = \mathbf{I} - \frac{S}{1+2SK}\mathbf{a}\mathbf{a}^H \tag{11.37}$$

so that

$$\sum_{t=1}^{N} \mathbf{x}^H(t)\mathbf{R}^{-1}\mathbf{x}(t) = N\,\mathrm{Tr}\{\hat{\mathbf{R}}\} - \frac{NS}{2SK+1}\mathbf{a}^H\hat{\mathbf{R}}\mathbf{a} \tag{11.38}$$

where $\mathrm{Tr}\{\cdot\}$ denotes the trace of a matrix and

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}(n)\mathbf{x}^H(n) \tag{11.39}$$

is the sample covariance matrix. To compute the determinant of $\mathbf{R}$ we use the rule $|\mathbf{I}_M + \mathbf{A}\mathbf{A}^H| = |\mathbf{I}_N + \mathbf{A}^H\mathbf{A}|$ for an $M \times N$ matrix $\mathbf{A}$ [Lüt96]. This gives

$$|\mathbf{R}| = |\mathbf{I}_{2K} + S\mathbf{a}\mathbf{a}^H| = |1 + S\mathbf{a}^H\mathbf{a}| = 2SK + 1 \tag{11.40}$$

The conditional density $p(\mathbf{X}|u)$ then reduces to

$$p(\mathbf{X}|u) = \frac{\exp\left[-N\mathrm{Tr}(\hat{\mathbf{R}})\right]}{\pi^{2KN}(2SK+1)^N} \exp\left[\frac{2SKN}{2SK+1}V(u)\right], \tag{11.41}$$

where

$$V(u) = \frac{\mathbf{a}^H(u)\hat{\mathbf{R}}\mathbf{a}(u)}{\mathbf{a}^H(u)\mathbf{a}(u)} = \frac{\mathbf{a}^H(u)\hat{\mathbf{R}}\mathbf{a}(u)}{2K} \tag{11.42}$$

is the conventional beamformer spectrum. The only quantity in the posterior PDF that depends on u is $V(u)$. Therefore, the ML and MAP estimators are identical to a conventional beamforming estimator in this case. Combining (11.35) and (11.41) we find that the MMSE estimator is given by

$$\hat{u}_{\mathrm{MMSE}} = \frac{\displaystyle\int_U u \exp\left[\frac{2SKN}{2SK+1}V(u)\right] du}{\displaystyle\int_U \exp\left[\frac{2SKN}{2SK+1}V(u)\right] du} \tag{11.43}$$

The MMSE estimator is thus given by the center of gravity of $\exp\left[\frac{2SKN}{2SK+1}V(u)\right]$. Notice that the MAP estimator is independent of S, and that the dependence of the MMSE estimator on S is weak if $2SK \gg 1$. The assumption that S is known is therefore not particularly restrictive for the single signal case.

Performance Evaluation

Next, we investigate the performances of the MMSE and MAP estimators, and compare them with the WWB and CRB. First, we show in a simulation example that the local MSE of the MAP and MMSE estimators is approximately independent of u. This was the justification for comparing classical and Bayesian bounds in the previous chapter. Figure 11.4 shows the local MSE for the MAP and MMSE estimators versus u. To avoid the edge effects described previously, the normalized distance between the sensors in a subarray was $d_n = 0.4$. Clearly, the local MSE is independent of u. It should be noted that it is not true that the local MSE is independent of u for very low SNR. As an example of this, consider a case where the SNR is so low that $\hat{u}$ is uniformly distributed between -1 and 1 irrespective of what the actual true value of u is. Now, if the true value is $u = 0$ we have that $\mathrm{E}[(\hat{u} - 0)^2] = E[\hat{u}^2] = 1/3$. If, however, $u = 1$ we have that $\mathrm{E}[(\hat{u} - 1)^2] = E[\hat{u}^2] + 1 = 4/3$. Therefore, in this case, the local MSE cannot be equal to the global MSE. The requirement for the local MSE to be independent of u is that the SNR is sufficiently high so that the estimates come from the mainlobe or its nearest grating lobes. If this is not the case, the estimates obtained are useless anyway, so this is not considered as a practical problem.

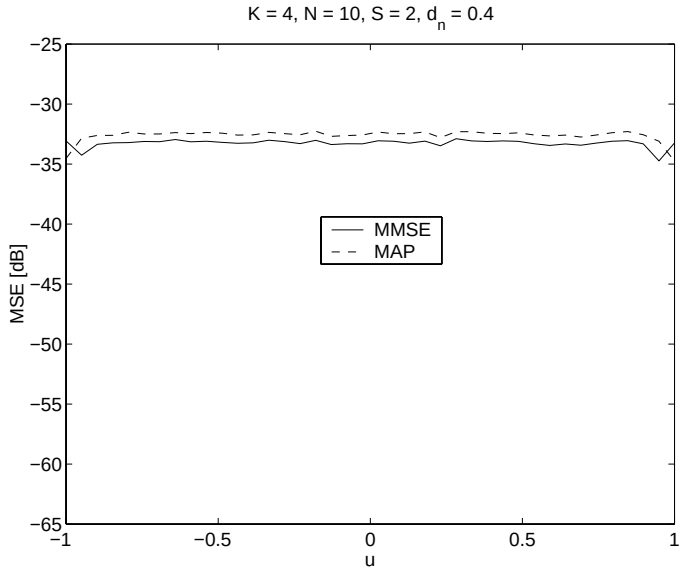

Figure 11.4: MSE of MAP and MMSE estimators vs u.

Figure 11.5 shows the MSE obtained by Monte Carlo simulations as a function of the distance D_n between the phase centers of the two subarrays.

For small subarray separation, it can be seen that CRB and WWB both are tight bounds. Beyond the ambiguity threshold however, there is gap between the WWB and the MSE of the MMSE estimator, which is the greatest lower bound. The WWB is thus not a tight bound in the ambiguity region. Comparing the MSE of the MAP and MMSE estimators we see that they have the same accuracy in the ambiguity free region, and that the difference in the ambiguity region is small. The location of the ambiguity threshold is the same for the MAP and MMSE estimators. The performance of the MAP estimator is thus very close to the performance of the optimal MMSE estimator in the region where a working system would be desired to operate. Since the MAP estimator is equivalent to a conventional beamformer in this case, we conclude that the beamformer, for all practical purposes, is the optimal estimator for this problem.

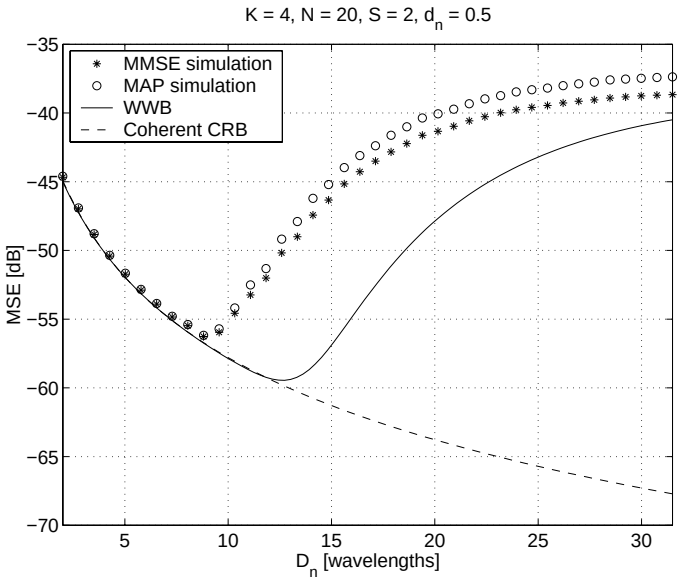

Figure 11.5: MSE of MAP and MMSE as a function of the distance between the subarray phase centers compared with WWB and CRB

Approximate MSE of the MAP Estimator

We have seen that the WWB is not a tight bound. The MSE of the MMSE estimator is a tight bound, but to the author's knowledge there exists no closed form expression for its computation. It would be desirable to have a simple, closed form expression for this MSE that could provide insights into array design trade-offs without having to resort to time-consuming Monte

Carlo simulations. We have already noted that the MAP estimator has essentially the same MSE as the MMSE estimator in the region where a working system would be designed to operate. Next, we derive a simple approximation for the MSE of the MAP estimator that accurately predicts the ultimate estimation accuracy that can be achieved with this class of arrays. Although not strictly a lower bound, we will show by simulation that this approximation is accurate enough as a tool for analyzing the fundamental limitations in DOA estimation with separated subarrays.

Derivation of the MSE

The approach we take is similar to the one reported in [HL00], where an approximate expression for the ML MSE has been derived heuristically for the problem of Doppler frequency estimation of a coherent pulse train. Since the MAP estimator is identical in form to the ML estimator if we assume a uniform prior PDF, this approximation can be applied also to the MAP estimator.

One way to interpret the array beampattern in Figure 11.2, is that the MAP (ML, beamformer) criterion function for coherent estimation is modulated by the criterion function for incoherent estimation. The peak of the "incoherent" envelope, or subarray pattern, will determine which of the narrower "coherent" peaks, or grating lobes, that will produce the MAP estimate. Ambiguous estimates will occur when fluctuations in the position of the peak of the subarray pattern cause the global maximum in the overall beampattern to appear at a grating lobe. For large number of snapshots or high SNR, the peak of the subarray pattern is assumed to be distributed according to

$$p(\hat{u}|u_0) = \frac{1}{\sqrt{2\pi B_I}} \exp\left[-\frac{1}{2B_I}(\hat{u}-u_0)^2\right] \tag{11.44}$$

This is motivated by general ML theory which states that the ML estimator is asymptotically Gaussian distributed with mean equal to the true value and variance equal to the CRB [Tre68]. Since ML, MAP and beamforming all are equivalent estimators in the present problem, this is a reasonable assumption. The probability of choosing the k-th peak in the array beampattern can then be expressed as

$$P_k = \int_{u_0+(2k-1)/2D_n}^{u_0+(2k+1)/2D_n} \frac{1}{\sqrt{2\pi B_I}} \exp\left[-\frac{1}{2B_I}(u-u_0)^2\right] du \tag{11.45}$$

where $1/D_n$ is the distance between the grating lobes. The numbering of k is such that $k = 0$ corresponds to the true peak in the array beampattern,

$k = 1$ corresponds to the first grating lobe to the right of the true peak, etc. These probabilities can be expressed in terms of the complimentary error function according to

$$P_k = \text{erfc}\left(\frac{(2k-1)}{2D_n\sqrt{B_I}}\right) - \text{erfc}\left(\frac{(2k+1)}{2D_n\sqrt{B_I}}\right) \tag{11.46}$$

where

$$\text{erfc}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \tag{11.47}$$

If the true peak is chosen, the MSE is given by the coherent CRB. If the k-th grating lobe is chosen, the contribution to the MSE is approximately $(k/D_n)^2$. The local MSE is then approximated by a sum of all MSE contributions, weighted by their respective probabilities of occurrence

$$\text{MSE}(\hat{u}) \approx P_0 B_C + \frac{2}{D_n^2} \sum_{k=1}^{M} k^2 P_k, \tag{11.48}$$

where M is the number of grating lobes included in the approximation. Since nothing depends on u in (11.48), averaging the local MSE over the prior PDF has no effect. The approximation of the local MSE in (11.48) is therefore also a valid approximation of the global MSE.

Interpretation of the MAP MSE

The approximation (11.48) provides valuable insights into the nature of the ambiguity effects on the MSE. Figure 11.7 shows the individual terms in (11.48) as a function of the subarray separation. For comparison purposes, the coherent CRB and the total MSE using all terms in (11.48) is included. It can be seen that the contribution of the mainlobe follows essentially the coherent CRB. For large separation, the mainlobe contribution starts to deviate from the CRB. This is due to that the probability of mainlobe estimates becomes appreciably below one. For small subarray separation, there is a good match between the total MSE and the CRB. In this region, there is no contribution from the grating lobes. At the threshold, the first grating lobe starts to give significant contribution and, at wider separation, also the second grating lobe begins to produce ambiguous estimates.

Further Simplification of the MAP MSE

We are interested in finding a good approximation of the MSE in the vicinity of the threshold. From Figure 11.7 we conclude that only the mainlobe and

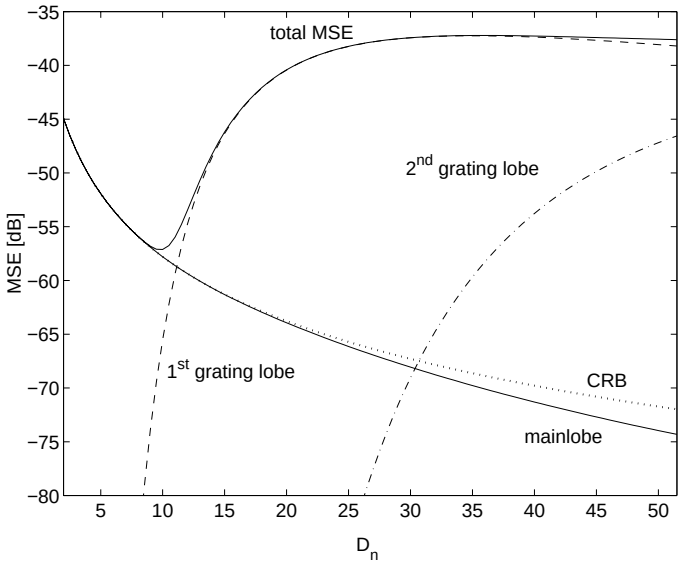

Figure 11.6: MSE contributions from the mainlobe and different grating lobes. $K = 4,\ N = 20,\ d_n = 0.5,\ S = 2$

the first grating lobe has significant contribution to the total MSE at the threshold. We also note that $P_0 \approx 1$ in this region. These observations together with the following approximation [Tre68]

$$\text{erfc}(x) \approx \frac{1}{\sqrt{2\pi}x} e^{-x^2/2} \quad x > 0, \tag{11.49}$$

leads to the following expression

$$\text{MSE}(\hat{u}) \approx B_C + \sqrt{\frac{8B_I}{\pi}} \frac{1}{D_n} \exp\left[-\frac{1}{8B_I D_n^2}\right] \tag{11.50}$$

where B_C and B_I are given in (11.12). Figure 11.7 shows the approximations in eq. (11.48) and (11.50) labeled Approximation 1 and Approximation 2 respectively. Clearly, the further approximations that took us to (11.50) were reasonable, since there is no visible difference between the two approximations except at large subarray separations. If the approximations are to be used for determining the ambiguity threshold, the simpler approximation in (11.50) is just as accurate as the one in (11.48). The ambiguity threshold can easily be obtained by finding the minimum of (11.50). Included in the plot is also the result from Monte Carlo simulation (also shown in Figure 11.5) of the MAP MSE. The MSE approximations derived are very good at predicting the MSE of the MAP estimator. Since the MAP MSE is very close

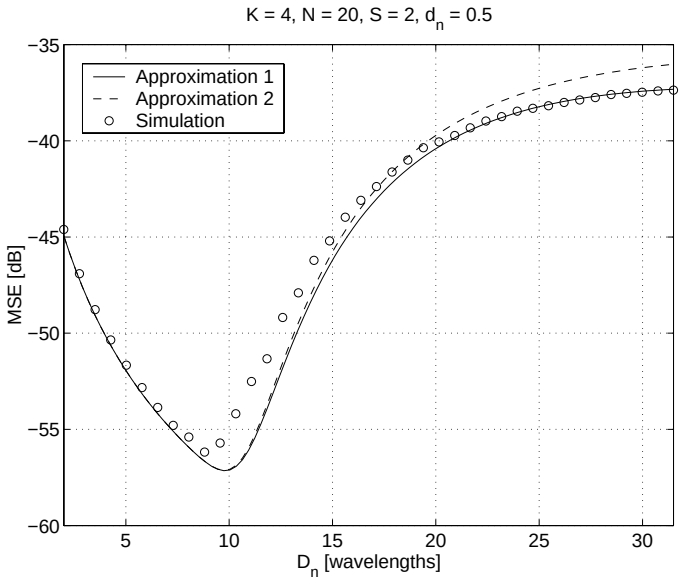

Figure 11.7: Approximations of MAP MSE compared with simulations

to the MMSE, we conclude that the approximations derived can be used to predict the ultimate estimation accuracy that can be achieved with this class of arrays.

As a last comparison, Figure 11.8 shows the MSE as a function of SNR for a fixed subarray separation. One can discern different SNR regions where the estimates have different nature. At high SNR, there are no unresolved ambiguities and the coherent CRB is reached. There is an ambiguity threshold also in this plot. Below this threshold the MSE increases to the incoherent CRB. At low SNR, only the envelope of the full array beampattern gives useful information. At very low SNR, the data provide no information and the MSE reaches the variance of the prior PDF. In this region, the MSE approximation is not good since the large estimation errors are not caused by grating lobes. The observant reader notes that points for the MMSE estimator are missing at high SNR. This is due to the extremely high dynamic range of the posterior PDF. Double precision in the computations was not sufficient to implement the MMSE estimator at these SNR values.

11.1.5 Conclusions

We have investigated the fundamental limitations in DOA estimation with a sensor array consisting of two spatially separated subarrays. Particular attention was paid to ambiguity errors caused by grating lobes in the ar-

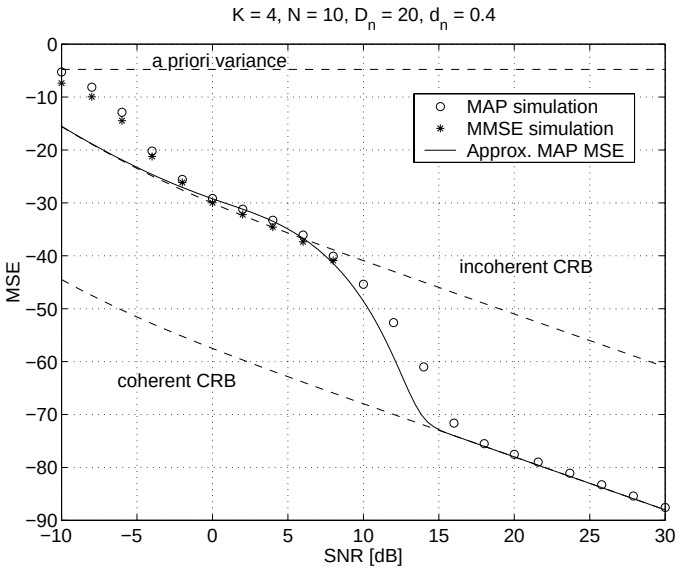

Figure 11.8: MSE as a function of SNR

ray beampattern. It was shown by simulation that the beamformer has essentially the same performance as the optimal MMSE estimator. A simple approximation of the beamformer's MSE that takes ambiguity errors into account was derived. This approximation provides valuable insights into how different array design parameters affect the attainable estimation accuracy.

11A Derivation of CRB for separated subarrays

The CRB based on the Gaussian signal model has been derived in [SN90b]. For the case of a single signal this reduces to

$$B_C = \frac{1}{2NS^2\mathbf{d}^H\mathbf{P}_{\mathbf{a}}^{\perp}\mathbf{d}\mathbf{a}^H\mathbf{R}^{-1}\mathbf{a}} \tag{11A.1}$$

where

$$\mathbf{d} = \frac{\partial\mathbf{a}(\theta)}{\partial u} \qquad \mathbf{P}_{\mathbf{a}}^{\perp} = \mathbf{I} - \mathbf{a}\mathbf{a}^H/\mathbf{a}^H\mathbf{a} = \mathbf{I} - \frac{1}{2K}\mathbf{a}\mathbf{a}^H \tag{11A.2}$$

Taking the derivative of the steering vector in (11.6) w.r.t. θ gives

$$\mathbf{d} = -j2\pi\mathbf{b} \odot \mathbf{a} \tag{11A.3}$$

where $\mathbf{b}\odot\mathbf{a}$ denotes element-wise multiplication of the elements in the vectors $\mathbf{b}$ and $\mathbf{a}$ and

$$\mathbf{b} \triangleq \left[\begin{array}{ccccccccc} 0 & d_n & \ldots & (K-1)d_n & D_n & D_n+d_n & \ldots & D_n+(K-1)d_n \end{array}\right]^T \tag{11A.4}$$

Denoting the elements in $\mathbf{b}$ by b_k we have

$$\begin{aligned} \mathbf{d}^H\mathbf{P}_{\mathbf{a}}^{\perp}\mathbf{d} &= 4\pi^2\left(\mathbf{b}^T\odot\mathbf{a}^H\right)(\mathbf{b}\odot\mathbf{a}) - \frac{2\pi^2}{K}\left(\mathbf{b}^T\odot\mathbf{a}^H\right)\mathbf{a}\mathbf{a}^H(\mathbf{b}\odot\mathbf{a}) \\ &= 4\pi\sum_{k=1}^{2K}b_k^2 - \frac{2\pi^2}{K}\left(\sum_{k=1}^{2K}b_k\right)^2 = \frac{2\pi^2}{3}K\left(d_n^2\left(K^2-1\right)+3D_n^2\right) \end{aligned} \tag{11A.5}$$

Applying the matrix inversion lemma to (11.7) gives

$$\mathbf{R}^{-1} = \mathbf{I} - \frac{S}{1+2SK}\mathbf{a}\mathbf{a}^H \tag{11A.6}$$

so that

$$\mathbf{a}^H\mathbf{R}^{-1}\mathbf{a} = 2K - \frac{S}{1+2SK}\mathbf{a}^H\mathbf{a}\mathbf{a}^H\mathbf{a} = \frac{2K}{1+2SK} \tag{11A.7}$$

Substituting (11A.5) and (11A.7) into (11A.1) then gives

$$B_C = \frac{3(2SK+1)}{8\pi^2NS^2K^2\left(d_n^2\left(K^2-1\right)+3D_n^2\right)} \tag{11A.8}$$

If $2KS \gg 1$ and $K^2 \gg 1$ this reduces to

$$B_C \approx \frac{3}{4\pi^2NKS\left(d_n^2K^2+3D_n^2\right)} \tag{11A.9}$$

The incoherent CRB, B_I, is given by the CRB for one subarray divided by two, since it is assumed that two independent estimates are obtained from the two subarrays. Thus,

$$B_I = \frac{1}{4NS^2\mathbf{d}_1^H\mathbf{P}_{\mathbf{a}_1}^{\perp}\mathbf{d}_1\mathbf{a}_1^H\mathbf{R}^{-1}\mathbf{a}_1}, \tag{11A.10}$$

where

$$\mathbf{d}_1 = \frac{\partial\mathbf{a}_1(\theta)}{\partial u} = -j2\pi\mathbf{b}_1\odot\mathbf{a}_1, \tag{11A.11}$$

and

$$\mathbf{b}_1 = \left[\begin{array}{cccc} 0 & d_n & \ldots & (K-1)d_n \end{array} \right]. \tag{11A.12}$$

In a similar fashion as for the coherent CRB we obtain

$$\mathbf{d}_1^H \mathbf{P}_{\mathbf{a}_1}^{\perp} \mathbf{d}_1 \;=\; 4\pi \sum_{k=1}^{K} b_{1k}^2 - \frac{2\pi^2}{K} \left(\sum_{k=1}^{K} b_{1k} \right)^2 = \frac{\pi^2 d_n^2 K(K^2-1)}{3} \tag{11A.13}$$

and

$$\mathbf{a}_1^H \mathbf{R}^{-1} \mathbf{a}_1 = \frac{K}{1+KS} \tag{11A.14}$$

so that

$$B_I = \frac{3(SK+1)}{4\pi^2 d_n^2 N S^2 K^2 \left(K^2-1\right)} \tag{11A.15}$$

If $2KS \gg 1$ and $K^2 \gg 1$ this reduces to

$$B_I \approx \frac{3}{4\pi^2 d_n^2 N S K^3} \tag{11A.16}$$

11.2 Optimization of Element Positions in Sparse Arrays

In this example we propose a method for optimizing the element positions in sparse linear arrays. Due to their high sidelobes, sparse arrays are plagued by the threshold effect. It is therefore important to acknowledge this effect in any sparse array design. We propose herein a novel criterion based on the SNR threshold in the Weiss-Weinstein bound (WWB).

11.2.1 Introduction

The DOA estimation accuracy depends to a great extent on the array size. Large arrays can thus provide very accurate estimates. DOA estimation for arrays with many elements are, however, expensive to implement, both in terms of receiver hardware and computational complexity.

For non-ambiguous DOA estimation with ULAs, the inter-element spacings should not exceed half a wavelength of the impinging wavefronts. In sparse arrays, elements are spaced further apart in order to obtain a large aperture with few elements. Sparse arrays thus have the potential of very accurate DOA estimation at a low cost. The price paid is the risk of obtaining ambiguous estimates, caused by grating lobes in the array beampattern. To reduce such grating lobes, non-uniform element spacing is employed. An important problem is then to determine which element positions yield the most accurate DOA estimates.

Different approaches to optimizing the element positions with respect to DOA estimation accuracy have been taken in the literature. In [AGGS96, GB97], the element positions of non-uniform linear arrays (NULAs) were optimized by minimization of the CRB. A problem with this approach is that the CRB is a local bound that does not take into account large estimation errors caused by near ambiguities. For the single signal problem this means that only the curvature of the mainlobe is considered; high sidelobes have no effect. At low SNR, these sidelobes may cause large estimation errors, rendering the CRB a far too optimistic bound in this case.

Various approaches have been proposed to account for near ambiguities. In [VE99] the mainlobe area was minimized subject to a peak sidelobe constraint and in [AS00] competitive criteria involving maximum aperture and identifiability were considered. Although these approaches are intuitively appealing, there is no explicit connection between these ambiguity/aperture trade-offs and the resulting mean square estimation error.

In this section, another approach is taken. A lower bound on the mean

square estimation error that takes ambiguity errors into account is used to optimize the element positions of a NULA with fixed aperture. The bound used is the Weiss-Weinstein Bound (WWB), which was first presented in [WW85] and subsequently applied to DOA estimation in e.g. [Ver87, DeL93, NV94, ZS98, AE00].

11.2.2 Problem Formulation

Consider a linear array of K sensors receiving a single planar wavefront from the DOA θ measured relative to the array boresight. For mathematical convenience, the estimation of $u \triangleq \sin\theta$ is considered. The element positions, denoted by $d_k, k = 1, \ldots, K$ are normalized by the standard spacing $\lambda/2$ where λ is the wavelength, i.e. $d_k = 2\tilde{d}_k/\lambda$ where $\tilde{d}_k$ is the physical distance. In the sequel, different linear array geometries, keeping the array length D (normalized by $\lambda/2$) fixed, will be studied. Without loss of generality, the end elements d_1 and d_K are fixed at 0 and D respectively. Assuming an ideal array with omnidirectional elements, the array output at time t can be modeled by the $K \times 1$ complex vector

$$\mathbf{x}(t) = \mathbf{a}(u)s(t) + \mathbf{n}(t), \qquad t = 1, \ldots, N \tag{11A.17}$$

where

$$\mathbf{a}(u) = \begin{bmatrix} 1 & e^{-j\pi d_2 u} & \ldots & e^{-j\pi d_{K-1} u} & e^{-j\pi D u} \end{bmatrix}^T$$

is the $K \times 1$ array steering vector. Furthermore, $s(t)$ denotes the impinging signal at baseband, $\mathbf{n}(t)$ is an additive noise term and N denotes the number of temporal snapshots. The signal $s(t)$ and noises $\mathbf{n}(t)$ are assumed independent and are modeled as white (spatially and temporally), zero mean, circular complex Gaussian random variables with second order moments

$$\mathrm{E}\left[|s(t)|^2\right] = \mathrm{SNR} \quad \text{and} \quad \mathrm{E}\left[\mathbf{n}(t)\mathbf{n}^H(t)\right] = \mathbf{I}. \tag{11A.18}$$

The signal variance is thus equal to the SNR per space-time sample since the noise variance has been normalized to unity. The SNR is further assumed to be known.

The problem considered in this example is, given the noisy observations $\mathbf{x}(t)$, $t = 1, \ldots, N$, to determine the element positions $d_2, \ldots, d_{K-1}$ that maximize the DOA estimation performance. A crucial issue that is addressed herein is how to define DOA estimation performance when the estimation is prone to ambiguities.

11.2.3 Optimization Method

The DOA estimation performance is often evaluated by means of the CRB. This bound is relatively easy to compute but is a local bound that does not take into account large errors that may be caused by near ambiguities. Various global bounds have been proposed in the literature. These are more tedious to compute but, on the other hand, they provide insights into how ambiguity errors affect the overall estimation error. One such bound is the Weiss-Weinstein Bound (WWB) [WW85]. This is a lower bound on the Mean Square Error (MSE) that rests on the Bayesian framework of estimation. This means that the parameter of interest is considered to be a random variable with known prior distribution. Throughout this section, a uniform distribution on $[-1, 1]$ is assigned to u. For details concerning the computation of the WWB for DOA estimation, see [DeL93].

To illustrate the difference between the WWB and the CRB, Figure 11.9 shows the CRB and WWB as a function of SNR for a particular NULA with 8 elements[3]. At high SNR, the WWB and CRB coincide since, in this region, ambiguous estimates do not occur. Below a certain SNR threshold the WWB increases rapidly. At this threshold, ambiguous estimates from grating lobes begin to yield contribution to the total MSE which is comparable to that of the mainlobe. This threshold effect is not captured by the CRB.

There are two main factors to consider when deciding upon an optimization criterion. On the one hand, the elements should be spread out as much as possible in order to get the lowest possible CRB. This implies good accuracy at high SNR. On the other hand, such a design will have a high SNR threshold. Since the array length is fixed, the difference in the CRB for different arrays will be relatively small. Therefore, we focus mainly on the SNR threshold.

The basic ideas behind the optimization procedure are as follows:

1. Generate a large number of different arrays with random element positions.
2. Compute the WWB as a function of SNR for each array.
3. Identify a reduced set of arrays with the lowest SNR thresholds.
4. The element positions of these arrays are used as starting points in a numerical optimization routine to improve the best arrays from the previous step.

[3]Since the WWB is a Bayesian bound, a direct comparison with the CRB is not meaningful in general. However, according to the discussion in Section 11.1.3, such a comparison can be justified in this case.

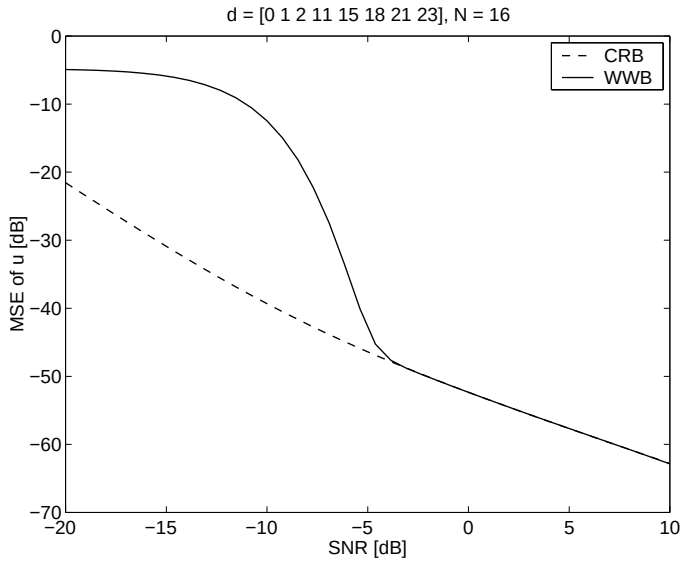

Figure 11.9: CRB and WWB as a function of SNR for a NULA.

5. The optimal element positions are then taken from the best array after the numerical optimization.

The SNR threshold can be defined in different ways, e.g. where the WWB exceeds the CRB by a certain amount or the maximum of the second derivative of the WWB curve. Unfortunately, we do not have an analytical expressions for the SNR threshold available. Furthermore, importance was attached to analyzing as many random arrays as possible within a limited computing time. Therefore, a somewhat simplified procedure was implemented:

- The WWB as a function of SNR was computed for 10^3 different arrays. The array with the lowest SNR threshold of these arrays was identified simply by looking at the plots of the WWB versus SNR. This array had a threshold at about SNR = -5 dB. Then, the WWB at SNR = -5 dB was computed for 10^6 random arrays.
- The 10 arrays with the lowest WWB at SNR = - 5 dB were selected for numerical optimization.
- The element positions of these arrays were used as starting points when minimizing the WWB with respect to element positions at SNR = -5 dB. The "fminsearch" routine in Matlab's Optimization Toolbox was used for this purpose.
- The array with the lowest WWB after the numerical optimization was then considered to be the optimal array.

A side-effect of this pragmatic approach is that we actually minimize the WWB for a fixed SNR and not the SNR threshold itself. However, since this fixed SNR is close to the threshold of the "best" array, we consider this an almost equivalent approach. In a way, it is a desired side-effect since we assure that we obtain also a low MSE; not just a low threshold.

There is no guarantee that the global optimum is found with this procedure. If a very large number of random arrays are generated, however, it is likely that the obtained solution is "sufficiently optimal" for any practical purposes.

The optimization procedure was evaluated by generating 10^6 eight-element linear arrays with random element positions and $D = 23, N = 16$. The element positions were generated according to a uniform distribution on $[0, D]$. Figure 11.10 shows the WWB as a function of SNR for the arrays which had the lowest and highest WWB at SNR = -5 dB. In order to show the

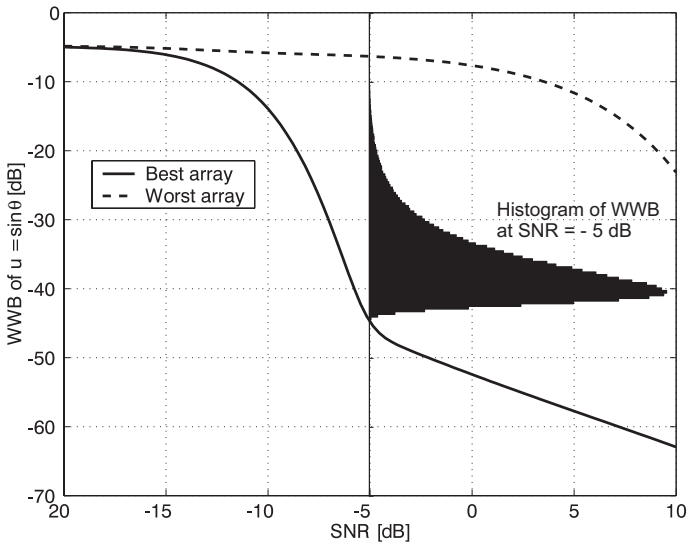

Figure 11.10: WWB for the best and worst array and a histogram of WWB at SNR = -5 dB.

statistical nature of the WWB of the randomly generated arrays, there is also a histogram of the WWB at SNR = -5 dB for all the arrays in the plot. The histogram has been rotated 90° compared to the standard orientation of a histogram. It can be seen from the figure that the difference between the WWB for the best and the worst array is quite large. The positions of the array elements thus have a great influence on the attainable estimation performance. The element positions of the 10 best arrays were then used as starting points in a numerical optimization routine to minimize the MSE

at SNR = -5 dB. Finally, the optimal element positions are taken from the array with the lowest MSE at this SNR. The numerical optimization reduced the minimum MSE from -44.8 dB to -45.2 dB.

Hitherto, the element positions were considered as continuous variables. This implies an infinite number of possible arrays with different element positions. On the other hand, constraining the element positions to a discrete grid leads to a finite number of possible different arrays. Therefore, it should be possible to compute the WWB for all these possible arrays if the number of grid points and array elements are not too large. A common approach is to start with a ULA with $\lambda/2$ element spacing and the required length. Then, a given number of elements are removed from the full array in order to produce the sparse array. These arrays are often called *thinned arrays*. In the present example, the two end elements are fixed. Thus, there are 22 element positions to choose 6 positions from. The number of different ways to pick 6 elements out of 22 is equal to $\binom{22}{6} = 74\,613$. This is a reasonable number of arrays for being able to compute the WWB for all of these arrays on a standard PC. The WWB at SNR = -5 dB was computed for all these 74 613 arrays and the result is illustrated in Figure 11.11. The WWB vs

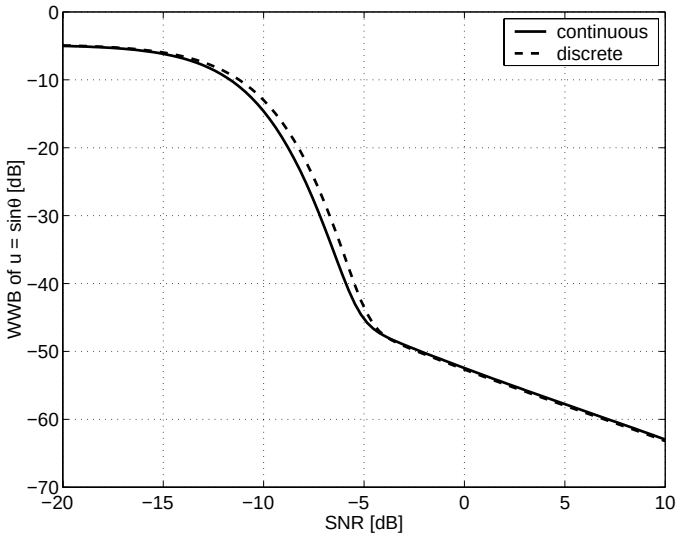

Figure 11.11: WWB for the best array using continuous and discrete element positions respectively.

SNR for the arrays with the lowest WWB at SNR = -5 dB using continuous and discrete positions respectively is shown. Clearly, the difference between the two is insignificant.

11.2.4 Comparison with other arrays

The arrays obtained from the optimization procedure described in the previous section were compared with a couple of other array configurations that have been studied in the literature. For such a comparison to be fair, all arrays had the same length and the same number of elements. A type of thinned array that has been widely studied is the so called minimum redundancy array [Mof68]. Another array configuration that also has been studied is two separated subarrays where each subarray is a ULA with $\lambda/2$ inter-element spacing, see e.g. [ZS98, BS01, ES02, Zat03] and Section 11.1. The array geometry that minimizes the CRB for one source and NULAs with fixed length is given by two point clusters at the array end points [GB97]. Due to mutual coupling effects and mechanical considerations, the element spacing cannot be too small. The separated subarrays configuration can thus be viewed as a realizable approximation of the CRB-optimal geometry. Figure 11.12 shows the element positions for the arrays under consideration. The beampatterns of the minimum redundancy array and the optimal thinned array are compared in Figure 11.13.

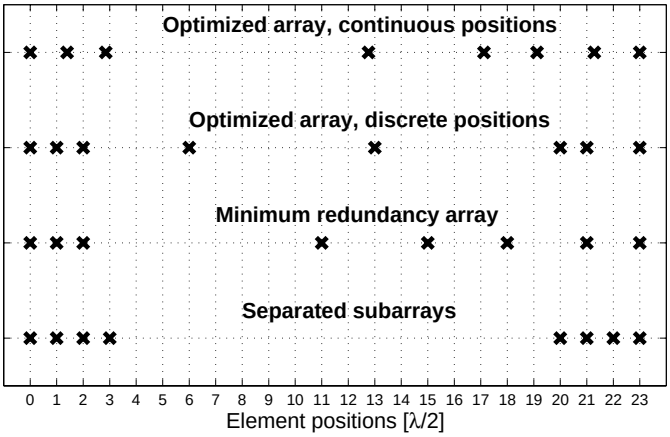

Figure 11.12: Element positions

In Figure 11.14, the WWB of the best thinned array obtained from the numerical optimization is compared with the WWB of a minimum redundancy array. There is practically no difference between the WWB for the two arrays. Recall that the difference between the WWB for the arrays obtained from minimization over continuous and discrete element positions respectively was very small. Therefore, it is concluded that the minimum redundancy array is nearly optimal under the criterion used in this example.

Figure 11.15 shows WWB vs SNR for the array obtained from optimiza-

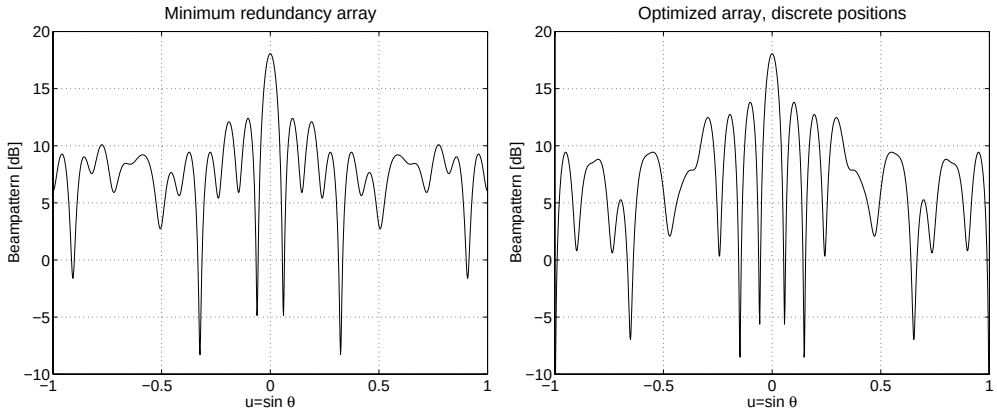

Figure 11.13: Array beampatterns

tion over continuous element positions and the separated subarray structure. The separated subarray structure has a considerably higher SNR threshold and somewhat lower WWB at high SNR as compared to the optimal array. This is expected, since the separated subarray structure has a narrower mainlobe but higher sidelobes, due to concentration of the elements near the array endpoints.

Common engineering practice suggests that low sidelobes are important for ambiguity-free DOA estimation. In order to investigate the adequacy of this in the current setting, Figure 11.16 displays a 2-D histogram of the WWB at SNR = -5 dB and the peak sidelobe in the beam pattern for each of the 10^6 arrays as a contour plot. Some interesting conclusions can be drawn from Figure 11.16. For most arrays, high peak sidelobe means high WWB. A few ridges can be discerned in the contour plot. These are probably due to the peak sidelobe being at different distances from the mainlobe. An ambiguous estimate from a sidelobe far from the mainlobe gives a greater contribution to the MSE than from a sidelobe close to the mainlobe. A interesting property appears if we scrutinize the lowest contour of the histogram. Apparently, if only the best arrays are considered, there seems to be no relationship between the peak sidelobe and the WWB, at least as long as the peak sidelobe does not exceed -3 dB (relative to the mainlobe). Inspection of the corresponding beampatterns revealed that for the arrays with low WWB and high peak sidelobe, the peak sidelobe is relatively close to the mainlobe. Obviously, as far as the MSE is concerned, a higher peak sidelobe can be tolerated if it is close to the mainlobe.

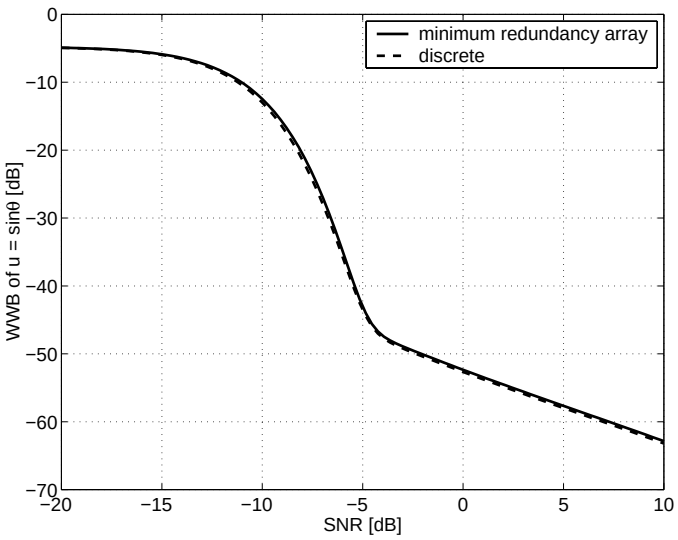

Figure 11.14: WWB for minimum redundancy array and the best thinned array.

11.2.5 Conclusions

A novel criterion for optimizing the element positions of sparse linear arrays has been presented. The criterion used was the ambiguity threshold of the Weiss-Weinstein Bound (WWB). This is a lower bound on the mean square DOA estimation error that takes into account large errors caused by ambiguities. An optimization procedure was implemented in order to find the array with lowest ambiguity threshold. The WWB for this array was compared with a minimum redundancy array and a separated subarrays structure. It was found that the optimal array and the minimum-redundancy array had nearly identical WWB.

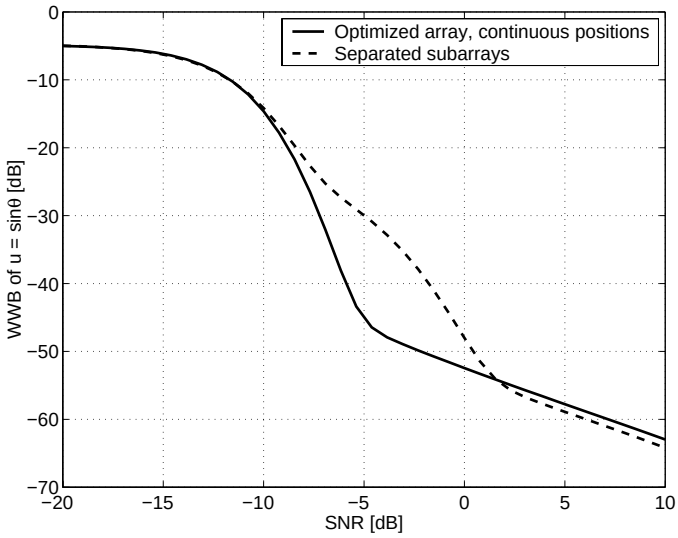

Figure 11.15: WWB for optimized array, continuous positions and separated subarrays respectively.

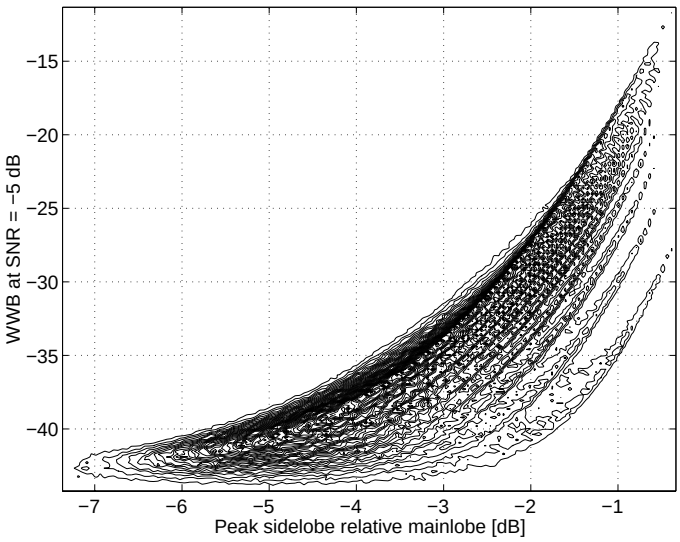

Figure 11.16: 2-D histogram of WWB at SNR = -5 dB and peak sidelobe, displayed as a contour plot.

Chapter 12

Conclusions

In the second part of this thesis we have developed a theory for predicting the threshold region performance of the ML DOA estimator. We treated the cases of deterministic and stochastic source signals for the single and multi-source estimation problems. Approximations of the MSE and probability of outlier were derived for the different cases of interest. Results of Monte Carlo simulations showed that these approximations could accurately predict the threshold region performance of the ML estimator.

The MSE approximation is based on a simple application of the total probability theorem. The difficulties arise when the probabilities of ambiguous estimates are to be computed. We found that these calculations can be treated in a common framework, since in all cases the calculations lead to computing the distribution of quadratic forms in complex Gaussian random variables. Closed form expressions for these probabilities were obtained, as well as fast and accurate approximations for cases when these expression were not practical to implement. As compared to Monte Carlo simulations, the analysis developed herein is a time-efficient alternative to evaluate the ML performance.

For the special case of a single stochastic source signal and a single snapshot, we showed that the ML estimator is not statistically efficient as $\mathrm{SNR} \to \infty$ due to the effect of outliers. Furthermore, results of Monte Carlo simulations indicated that this is also the case for multiple snapshots. However, in this case the inefficiency is due to the local errors not being small enough to reach the CRB.

We also presented two examples concerning DOA estimation with sparse arrays. In the first example we studied an interferometric-like system composed of two separated subarrays. We derived the Bayes-optimal estimator for this problem and compared it to other estimators as well as lower bounds

on the estimation performance. We found that the conventional beamformer, the ML, and the Bayesian MAP and MMSE estimators had essentially the same performance in this example. A simple expression for the MSE of the MAP estimator was then derived. This was shown to be in good agreement with results of Monte Carlo simulations.

In the second example we used the Weiss-Weinstein lower bound to optimize the element positions of sparse linear arrays. By minimizing the lower bound on the MSE close to the SNR threshold we obtained an array configuration that has high asymptotic estimation accuracy and yet is robust to ambiguity errors. The array obtained from the optimization was compared with a minimum redundancy array with the same length and the same number of elements. We found that these arrays have nearly identical Weiss-Weinstein lower bounds.

We have identified some topics for future research. Our observation that the stochastic ML estimator appears to be statistically inefficient for high SNR needs to be verified analytically. We have only showed it for the special case of a single source and a single snapshot. The theory for the multi-source stochastic ML estimator can be developed further. In particular, the probabilities associated with the concentrated likelihood function remains to be computed.

The quest for tight and computable lower bounds on estimation accuracy is still ongoing. It would be interesting if some connections between existing bounds and the approximations developed herein could be established. Maybe this could lead to new insights and improved bounds.

Finally, it would be interesting to apply the MSE approximations to the problem of optimizing the element positions in sparse arrays. In particular, an optimization criterion involving the capability of resolving two sources would be interesting to investigate.

Bibliography

[AE00] F. Athley and C. Engdahl. "Direction-of-Arrival Estimation Using Separated Subarrays". In *Proc. 34th Asilomar Conf. Signals, Syst., Comput.*, 2000.

[AGGS96] Y. I. Abramovich, D. A. Gray, A. Y. Gorokhov, and N. K. Spencer. "Comparison of DOA Estimation Performance for Various Types of Sparse Antenna Array Geometries". In *Proceedings of EUSIPCO-96*, 1996.

[AS00] Y. I. Abramovich and N. K. Spencer. "Design of Nonuniform Linear Antenna Array Geometry and Signal Processing Algorithm for DOA Estimation of Gaussian Sources". *Digital Signal Processing*, 10(4):340–54, October 2000.

[Asy98] Asym P. Tomaniac, a.k.a. P. Stoica. "Asymptomania?". *IEEE Signal Processing Magazine*, 15(1):16, Jan. 1998.

[Ath99] F. Athley. "Asymptotically Decoupled Angle-Frequency Estimation with Sensor Arrays". In *Proc. 33:rd Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, USA, 1999.

[Ath01a] F. Athley. "Angle and Frequency Estimation Using Sensor Arrays". Technical Report No. 370L (Lic. thesis), Chalmers University of Technology, Göteborg, Sweden, Feb. 2001.

[Ath01b] F. Athley. "Optimization of Element Positions for Direction Finding with Sparse Arrays". In *Proc. 11th IEEE Signal Processing Workshop on Statistical Signal Processing*, Singapore, 2001.

[Ath02a] F. Athley. "Performance Analysis of DOA Estimation in the Threshold Region". In *Proc. ICASSP 2002*, Orlando, FL, USA, 2002.

[Ath02b] F. Athley. "Threshold Region Performance of Deterministic Maximum Likelihood DOA Estimation of Multiple Sources". In *Proc. 36th Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, USA, 2002.

[Ath02c] F. Athley. "Threshold Region Performance of Maximum Likelihood DOA Estimation for a Single Source". In *Proc. 10th Annual Workshop on Adaptive Sensor Array Processing*, MIT Lincoln Lab., Lexington, MA, USA, 2002.

[AVE] F. Athley, M. Viberg, and J. Eriksson. "High-Resolution Space-Time Signal Processing for Radar". In Y. Hua, A. Gershman, and Q. Cheng, editors, *High-Resolution and Robust Signal Processing*. Dekker, New York, NY, USA. To be pubslihed.

[Bag89] E. J. Baghdady. "Hybrid Interferometry for High-Resolution DOA Measurement". In *Proc. IEEE National Aerospace Electronics Conf.*, 1989.

[Ban71] W. J. Bangs. *Array Processing with Generalized Beamformers*. PhD thesis, Yale University, New Haven, CT, 1971.

[Bar49] E. W. Barankin. "Locally Best Unbiased Estimates". *Ann. Math. Statist.*, 20:477–501, 1949.

[Bar88] D. K. Barton. *Modern Radar System Analysis*. Artech House, Norwood, 1988.

[BB00] S. Basu and Y. Bresler. "A Global Lower Bound on Parameter Estimation Error with Periodic Distortion Functions". *IEEE Transactions on Information Theory*, 46(3):1145–50, 2000.

[Bel95] K. L. Bell. *Performance Bounds in Parameter Estimation with Application to Bearing Estimation*. PhD thesis, George Mason University, Fairfax, VA, 1995.

[Ben93] G. R. Benitz. "Asymptotic Results for Maximum Likelihood Estimation with an Array of Sensors ". *IEEE Trans. on Information Theory*, 39(4):1374–1385, July 1993.

[BET96] K. L. Bell, Y. Ephraim, and H. L. Van Trees. "Explicit Ziv-Zakai Lower Bound for Bearing Estimation". *IEEE Transactions on Signal Processing*, 44(11):2810–2824, Nov. 1996.

[Böh84] J.F. Böhme. "Estimation of Source Parameters by Maximum Likelihood and Nonlinear Regression". In *Proc. ICASSP 84*, pages 7.3.1–7.3.4, 1984.

[Böh86] J. F. Böhme. "Estimation of Spectral Parameters of Correlated Signals in Wavefields". *Signal Processing*, 11:329–337, 1986.

[Böh95] J. F. Böhme. Statistical array signal processing of measured sonar and seismic data. In *Proc. SPIE, The International Society for Optical Engineering*, volume 2563, San Diego, 1995.

[Bom00] K. Boman. "Low-Angle Estimation: Models, Methods and Bounds". Technical Report Lic. Thesis, Uppsala University, 2000.

[Box54] G. E. P. Box. "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-way Classification". *Ann. Math. Statist.*, 25:290–302, 1954.

[Bra62] R. N. Bracewell. Radio astronomy techniques. In *Handbuch der Physik*, pages 42–129. Springer, Berlin, 1962.

[Bre61] L. E. Brennan. "Angular Accuracy of a Phased Array Radar". In *IRE Trans. on Antennas and Propagation*, pages 268–275, 1961.

[Bre78] J. W. Brewer. "Kronecker Products and Matrix Calculus in System Theory". *IEEE Trans. on Circuits and Systems*, 25(9):772–781, Sept. 1978.

[Bre88] Y. Bresler. "Maximum Likelihood Estimation of Linearly Structured Covariance with Application to Antenna Array Processing". In *Proc. 4th ASSP Workshop on Spectrum Estimation and Modeling*, pages 172–175, Minneapolis, MN, Aug. 1988.

[BS01] K. Boman and P. Stoica. "Low Angle Estimation: Models, Methods, and Bounds". *Digital Signal Processing*, 11:35–79, 2001.

[Cap69] J. Capon. "High Resolution Frequency Wave Number Spectrum Analysis". *Proc. IEEE*, 57:1408–1418, 1969.

[Car88] B. D. Carlson. "Covariance Matrix Estimation Errors and Diagonal Loading in Adaptive Arrays". *IEEE Trans. on Aerospace and Electronic Systems*, 24(4):397–401, July 1988.

[CF02] G. Cohen and J. M. Francos. "Least squares estimation of 2-D sinusoids in colored noise: asymptotic analysis ". *IEEE Trans. on Information Theory*, 48(8):2243–2252, Aug. 2002.

[CH96] Q. Cheng and Y. Hua. "Detection of Two-Dimensional Frequencies". In *Proc. ISSPA*, Bisbane, Australia, 1996.

[Cla93] M. P. Clark. "Cramer-Rao bounds for Two-Dimensional Deterministic Modal Analysis". In *Proc. 27th Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, USA, 1993.

[Cox73] H. Cox. "Resolving Power and Sensitivity to Mismatch of Optimum Array Processors". *Journal of the Acoustical Society of America*, 54(3):771–785, Sept. 1973.

[Cra46] H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, NJ., 1946.

[Dan54] H. E. Daniels. "Saddlepoint Approximations in Statistics". *Annals of Mathematical Statistics*, 25:631–650, 1954.

[DeL93] D. F. DeLong. "Use of the Weiss-Weinstein Bound to Compare the Direction-Finding Performance of Sparse Arrays". Technical Report TR-982, MIT Lincoln Lab., Aug. 1993.

[DL97] H. Dropkin and C. Ly. "Superresolution for Scanning Antenna". In *IEEE National Radar Conference*, Syracuse, NY, USA, 1997.

[DM84] D. E. Dudgeon and R. M. Merserau. *Multidimensional Digital Signal Processing*. Prentice-Hall, Englewood-Cliffs, NJ, 1984.

[DN02] A. Dogandžić and A. Nehorai. "Space-Time Fading Channel Estimation and Symbol Detection in Unknown Spatially Colored Noise". *IEEE Transactions on Signal Processing*, 50(3):457–474, March 2002.

[DS83] J.E. Dennis and R.B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice Hall, Englewood Cliffs, NJ., 1983.

[Eri01] J. Eriksson. Multiple signal detection under a false alarm constraint. In *Proc. ICASSP*, Salt Lake City, USA, 2001.

[Eri02] J. Eriksson. *On Detection and Estimation of Multiple Sources in Radar Array Processing*. PhD thesis, Chalmers University of Technology, Göteborg, Sweden, March 2002.

[ES02] C. Engdahl and P. Sunnergren. "Model-based Adaptive Detection and DOA Estimation Using Separated Sub-arrays ". In *Proc. 2002 IEEE Radar Conference*, Long Beach, CA, USA, 2002.

[Far92] A. Farina. *Antenna-Based Signal Processing Techniques for Radar Systems*. Artech House, Boston, 1992.

[Fis22] R. A. Fisher. "On the Mathematical Foundations of Theoretical Statistics". *Phil. Trans. Royal Soc.*, 222, 1922.

[Ful96] W. A. Fuller. *Introduction to Statistical Time Series*. Wiley, New York, 2nd edition, 1996.

[GB97] A. B. Gershman and J. F. Böhme. "A Note on Most Favorable Array Geometries for DOA Estimation and Array Interpolation". *IEEE Signal Processing Letters*, 4(8):232–5, August 1997.

[Goo63] N.R. Goodman. "Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution (An Introduction)". *Annals Math. Stat.*, Vol. 34:152–176, 1963.

[GP73] G. H. Golub and V. Pereyra. "The Differentiation of Pseudo-inverses and Nonlinear Least Squares Problems whose Variable Separate". *SIAM Journal of Numerical Analysis*, 10(2):413–432, April 1973.

[GR94] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, Inc., Boston, fifth edition, 1994.

[GS92] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes*. Oxford Science Publications, Oxford, second edition, 1992.

[Hay95] S. Haykin, editor. *Advances in Spectrum Analysis and Array Processing*, volume I-III. Prentice-Hall, 1991-1995.

[Hel86] C.W. Helstrom. "Calculating Error Probabilities for Intersymbol and Cochannel Interference". *IEEE Trans. Commun.*, COM-34(5):430–5, 1986.

[HGC] Y. Hua, A. Gershman, and Q. Cheng, editors. *High-Resolution and Robust Signal Processing*. Dekker, New York, NY, USA. To be pubslihed.

[HL00] S. D. Howard and P. Lavoie. "Analysis of SNR Threshold for Differential Doppler Frequency Measurement in Digital Receivers". In *Proc. ICASSP 2000*, Istanbul, Turkey, 2000.

[HLE93] S. Haykin, J. Litva, and T. J. Shepherd (Eds.). *Radar Array Processing*. Springer-Verlag, Berlin, 1993.

[HRKV92] S. Haykin, J. P. Reilly, V. Kezys, and E. Vertatschitsch. "Some Aspects of Array Signal Processing". *IEE Proceedings-F*, 139(1):1–26, February 1992.

[HS92] S. Haykin and A. Steinhardt, editors. *Adaptive Radar Detection and Estimation*. John Wiley & Sons, Inc, New York, 1992.

[Hua92] Y. Hua. "Estimating Two-Dimensional Frequencies by Matrix Enhancement and Matrix Pencil". *IEEE Trans. on Signal Processing*, 40(9):2267–2280, September 1992.

[Imh61] J. P. Imhof. "Computing the Distribution of Quadratic Forms in Normal Variables". *Biometrika*, 48(3/4):419–426, 1961.

[JD93] D. H. Johnson and D. E. Dudgeon. *Array Signal Processing. Concepts and Techniques*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1993.

[JK70] N. L. Johnson and S. Kotz. *Distributions in Statistics: Continuous Univariate Distributions - 2*. Wiley, New York, 1970.

[JR81] E. Jacobs and E. W. Ralston. "Ambiguity Resolution in Interferometry". *IEEE Trans. on Aerospace and Electronic Systems*, 17(6):766–780, Nov. 1981.

[Kau75] L. Kaufman. "A Variable Projection Method for Solving Separable Nonlinear Least Squares Problems". *BIT*, 15:49–57, 1975.

[Kay80] S. M. Kay. *Modern Spectral Estimation: Theory and Application*. Prentice-Hall, Englewood-Cliffs, NJ, 1980.

[Kay93] S. M. Kay. *Fundamentals of Statistical Signal Processing. Vol. 1 Estimation Theory*. Prentice Hall, Upper Saddle River, N.J., 1993.

[Kay98] S. M. Kay. *Fundamentals of Statistical Signal Processing. Vol. 2 Detection Theory.* Prentice Hall, Upper Saddle River, N.J., 1998.

[Kle98] R. Klemm. *Space-Time Adaptive Processing: Principles and Applications.* IEE, London, 1998.

[KM96] D. Kundu and A. Mitra. "Asymptotic Properties of the Least Squares Estimates of 2-D Exponential Signals". *Multidimensional Systems and Signal Processing*, 7:135–150, 1996.

[KN01] D. Kundu and S. Nandi. "On Asymptotic Properties of a Two Dimensional Frequency Estimator". *Communications in Statistics — Theory and Methods*, 30(8&9):1561–1577, July 2001.

[Kuo99] D. Kuonen. "Saddlepoint Approximations for Distributions of Quadratic Forms in Normal Variables". *Biometrika*, 86(4):929–935, 1999.

[KV96] H. Krim and M. Viberg. "Two Decades of Array Signal Processing Research: The Parametric Approach". *IEEE Signal Processing Magazine*, 13(4):67–94, July 1996.

[Leh83] E.L. Lehmann. *Theory of Point Estimation.* John Wiley & Sons, New York, 1983.

[LHSV95] J. Li, B. Halder, P. Stoica, and M. Viberg. "Computationally Efficient Angle Estimation for Signals with Known Waveforms". *IEEE Trans. on Signal Processing*, 43(9):2154–2163, Sep. 1995.

[Lju99] L. Ljung. *System Identification. Theory for the User.* Prentice Hall, Upper Saddle River, N.J., 2nd edition, 1999.

[LL66] Y. T. Lo and S. W. Lee. "A Study of Space-Tapered Arrays". *IEEE Trans. on Antennas and Propagation*, 14(1):22–30, Jan. 1966.

[LR93] J. Li and R. T. Compton, Jr. "Maximum Likelihood Angle Estimation for Signals with Known Waveforms". *IEEE Trans. on Signal Processing*, 41(9):2850–2862, Sep. 1993.

[LST93] D. A. Linebarger, I. H. Sudborough, and I . G. Tollis. "Difference Bases and Sparse Sensor Arrays". *IEEE Trans. on Information Theory*, 39(2):716–721, March 1993.

[LSZ96] J. Li, P. Stoica, and D. Zheng. "An Efficient Algorithm for Two-Dimensional Frequency Estimation". *Multidimensional Systems and Signal Processing*, 7:151–178, April 1996.

[Lüt96] H. Lütkepohl. *Handbook of Matrices*. Wiley, 1996.

[LvdVD98] A. N. Lemma, A. J. van der Veen, and E. F. Deprettere. "Joint Angle-Frequency Estimation Using Multi-Resolution ESPRIT". In *Proc. ICASSP 98*, 1998.

[LvdVD99] A. N. Lemma, A. J. van der Veen, and E. F. Deprettere. "Multiresolution ESPRIT Algorithm". *IEEE Trans. on Signal Processing*, 47(6):1722–1726, Jun. 1999.

[LZS93] J. Li, D. Zheng, and P. Stoica. "Angle and Waveform Estimation Via RELAX". *IEEE Trans. on Aerospace and Electronic Systems*, 33(3):1077–1087, July 1993.

[Mai94] R. J. Mailloux. *Phased Array Antenna Handbook*. Artech House, Boston, 1994.

[McC82] J. H. McClellan. "Multidimensional Spectral Estimation". *Proceedings of the IEEE*, 70(9):1029–1039, September 1982.

[Mes92] H. Messer. "Source Localization Performance and the Array Beampattern". *Signal Processing*, 28(2):163–181, Aug. 1992.

[MH71] R. J. McAulay and E. M. Hofstetter. "Barankin Bounds on Parameter Estimation". *IEEE Trans. on Information Theory*, 17(6):669–676, Nov. 1971.

[MKB79] K.V. Mardia, J.T. Kent, and J.M. Bibby. *Multivariate Analysis*. Academic Press, London, 1979.

[ML00] Y. Ma and T. J. Lim. "Bit Error Probability for MDPSK and NCFSK over Arbitrary Rician Fading Channels". *IEEE J. Select. Areas Commun.*, 18(11):2179–89, 2000.

[MLP02] Y. Ma, T. J. Lim, and S. Pasupathy. "Error Probability for Coherent and Differential PSK over Arbitrary Rician Fading Channels with Multiple Cochannel Interferers". *IEEE Trans. Commun.*, 50(3):429–41, 2002.

[MM80] R. A. Monzingo and T. W. Miller. *Introduction to Adaptive Arrays*. John Wiley and Sons, 1980.

[Mof68] A. T. Moffet. "Minimum-Redundancy Linear Arrays". *IEEE Trans. Antennas Propagat.*, AP-16(2):172–175, March 1968.

[MP92] A. M. Mathai and S. B. Provost. *Quadratic Forms in Random Variables : Theory and Applications*. Dekker, 1992.

[MS69] R. J. McAulay and L. P. Seidman. "A Useful Form of the Barankin Lower Bound and its Application to PPM Threshold Analysis". *IEEE Trans. on Information Theory*, 15(2):273–279, March 1969.

[MS02] A. Mitra and P. Stoica. "The Asymptotic Cramér-Rao Bound for 2-D Superimposed Exponential Signals". *Multidimensional Systems and Signal Processing*, 13:317–331, July 2002.

[Ngu94] H. Nguyen. *Robust Beamforming Using Direction of Arrival Estimates*. PhD thesis, George Mason University, Fairfax, VA, 1994.

[Nic87] U. Nickel. "Angular Superresolution with Phased Array Radar: A Review of Algorithms and Operational Constraints". *IEE Proceedings F*, 134(1):53–9, 1987.

[NRC91] F. E. Nathanson, J. P. Reilly, and M. N. Cohen. *Radar Design Principles: Signal Processing and the Environment*. McGraw-Hill, New York, 2nd edition, 1991.

[NV94] H. Nguyen and H. L. Van Trees. "Comparison of Performance Bounds for DOA Estimation". In *Seventh SP Workshop on Statistical Signal & Array Processing*, pages 313–316, 1994.

[OVK92] B. Ottersten, M. Viberg, and T. Kailath. "Analysis of Subspace Fitting and ML Techniques for Parameter Estimation from Sensor Array Data". *IEEE Trans. on SP*, SP-40:590–600, March 1992.

[OVSN93] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai. "Exact and Large Sample Maximum Likelihood Techniques for Parameter Estimation and Detection in Array Processing". In S. Haykin, J. Litva, and T. J. Shepherd, editors, *Radar Array Processing*. Springer-Verlag, Berlin, 1993.

[PR96] S. B. Provost and E. D. Rudiuk. "The Exact Distribution of Indefinite Quadratic Forms in Noncentral Normal Vectors". *Annals of the Institute of Statistical Mathematics*, 48(2):381–394, 1996.

[Pro95] J. G. Proakis. *Digital Communications*. McGraw-Hill, Inc., 3rd edition, 1995.

[Rao45] C.R. Rao. "Information and Accuracy Attainable in the Estimation of Statistical Parameters". *Bull. Calcutta Math. Soc.*, 37:81–91, 1945.

[Rap96] D. Raphaeli. "Distribution of Noncentral Indefinite Quadratic Forms in Complex Normal Variables". *IEEE Trans. Inform. Theory*, 42(3):1002–7, 1996.

[RB74] D. C. Rife and R. B. Boorstyn. "Single-Tone Parameter Estimation from Discrete-Time Observations". In *IEEE Trans. on Information Theory*, pages 591–598, 1974.

[Rei88] N. Reid. "Saddlepoint Methods and Statistical Inference". *Statistical Science*, 3(2):213–238, 1988.

[Ric80] S. O. Rice. "Distribution of Quadratic Forms in Normal Random Variables - Evaluation by Numerical Integration". *SIAM J. Sci. Stat. Comput.*, 1(4):438–448, Dec. 1980.

[Ric03] C. Richmond. "Mean-Squared Error Performance Prediction of Maximum-Likelihood Signal Parameter Estimation ". In *Proc. 11th Annual Workshop on Adaptive Sensor Array Processing*, MIT Lincoln Lab., Lexington, MA, USA, 2003.

[RPK86] R. Roy, A. Paulraj, and T. Kailath. "ESPRIT – A Subspace Rotation Approach to Estimation of Parameters of Cisoids in Noise". *IEEE Trans. on ASSP*, ASSP-34(4):1340–1342, October 1986.

[RW80] A. Ruhe and P.Å. Wedin. "Algorithms for Separable Nonlinear Least Squares Problems". *SIAM Review*, 22:318–327, July 1980.

[RW90] L. Råde and B. Westergren. *Beta. Mathematics Handbook.* Lund: Studentlitteratur; Bromley: Chartwell-Bratt, 2nd edition, 1990.

[RZ93] C. R. Rao and L. C. Zhao. "Asymptotic Behavior of Maximum Likelihood Estimates of Superimposed Exponential Signals". In *IEEE Trans. on Signal Processing*, pages 1461–1464, 1993.

[RZZ93] C. R. Rao, L. C. Zhao, and B. Zhou. "A Novel Algorithm for 2-Dimensional Frequency Estimation". In *Proc. 27th Asilomar*

Conf. on Signals, Systems and Computers, Pacific Grove, CA, USA, 1993.

[RZZ94] C. R. Rao, L. C. Zhao, and B. Zhou. "Maximum Likelihood Estimation of 2-D Superimposed Exponential Signals". *IEEE Trans. on Signal Processing*, 42(7):1795–1802, July 1994.

[Sch79] R.O. Schmidt. "Multiple Emitter Location and Signal Parameter Estimation". In *Proc. RADC Spectrum Estimation Workshop*, pages 243–258, Rome, NY, 1979.

[Shu83] R. H. Shumway. Replicated time-series regression: An approach to signal estimation and detection. In *D.R. Brillinger and P. R. Krishnaiah (eds), Handbook of Statistics*, pages 383–408. Elsevier Science Publishers B. V., Inc, 1983.

[Sko86] M. I. Skolnik. *Introduction to Radar Systems*. McGraw-Hill, New York, 2nd edition, 1986.

[Sko89] M. I. Skolnik. *Radar Handbook*. McGraw-Hill, New York, 2nd edition, 1989.

[SM97] P. Stoica and R. Moses. *Introduction to Spectral Analysis*. Prentice-Hall, Upper Saddle River, NJ, 1997.

[SN89] P. Stoica and A. Nehorai. "MUSIC, Maximum Likelihood, and Cramér-Rao Bound". *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 37(5):720–741, May 1989.

[SN90a] P. Stoica and A. Nehorai. "MUSIC, Maximum Likelihood, and Cramér-Rao Bound: Further Results and Comparisons". *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 38(12):2140–2150, Dec. 1990.

[SN90b] P. Stoica and A. Nehorai. "Performance Study of Conditional and Unconditional Direction-of-Arrival Estimation". *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 38(10):1783–1795, Oct. 1990.

[SN95] P. Stoica and A. Nehorai. "On the Concentrated Stochastic Likelihood Function in Array Signal Processing". *Circuits, Systems, and Signal Processing*, 14(5):669–674, 1995.

[SOVM96] P. Stoica, B. Ottersten, M. Viberg, and R. Moses. "Maximum Likelihood Array Processing for Stochastic Coherent Sources". *IEEE Trans. SP*, SP-44:96–105, Jan. 1996.

[SS89] P. Stoica and T. Söderstrom. "On Reparametrization of Loss Functions Used in Estimation and the Invariance Principle". *Signal Processing*, 17:383–387, Aug. 1989.

[SS92] P. Stoica and T. Söderström. "On Array Signal Processing in Spatially Correlated Noise Fields ". *IEEE Trans. on Circuits and Systems II: Analog and Digital Signal Processing*, 39(12):879–882, Dec. 1992.

[SS98] A. L. Swindlehurst and P. Stoica. "Maximum Likelihood Methods in Radar Array Processing". *Proceedings of the IEEE*, 86(2):421–441, Feb. 1998.

[SSM93] J. L. Sacchini, W. M. Steedly, and R. L. Moses. "Two-Dimensional Prony Modeling and Parameter Estimation". *IEEE Trans. on Signal Processing*, 41(11):3127–3137, November 1993.

[TK99] J. Tabrikian and J. L. Krolik. "Barankin Bounds for Source Localization in an Uncertain Ocean Environment". *IEEE Trans. on Signal Processing*, 47(11):2917–2927, Nov. 1999.

[Tre68] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I.* John Wiley and Sons, New York, 1968.

[Tre02] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part IV, Optimum Array Processing.* John Wiley and Sons, 2002.

[Tur60] G. L. Turin. "The Characteristic Function of Hermitian Quadratic Forms in Complex Normal Variables". *Biometrika*, 47(1/2):199–201, June 1960.

[VB88] B. D. Van Veen and K. M. Buckley. "Beamforming: A versatile approach to spatial filtering". *IEEE ASSP Magazine*, 5(2):4–24, 1988.

[VE99] M. Viberg and C. Engdahl. "Element position considerations for robust direction finding using sparse arrays". In *Proc.* 33^{rd} *Asilomar Conf. on Signals, Systems, and Computers*, 1999.

[Ver87] E. J. Vertatschitsch. *Linear Array Structures for Direction of Arrival Estimation.* PhD thesis, McMaster Univ., Hamilton, ON, Canada, 1987.

[VH91] E. J. Vertatschitsch and S. Haykin. "Impact of Linear Array Geometry on Direction-of-Arrival Estimation for a Single Source". *IEEE Trans. on Antennas and Propagation*, 39(5):576–84, 1991.

[Vib93] M. Viberg. "Sensitivity of Parametric Direction Finding to Colored Noise Fields and Undermodeling". *Signal Processing*, 34(2):207–22, 1993.

[Vib95] M. Viberg. "Performance Analysis of Direction Finding with Large Arrays and Finite Data". *IEEE Trans. on Signal Processing*, 43(2):469–477, February 1995.

[VO91] M. Viberg and B. Ottersten. "Sensor Array Processing Based on Subspace Fitting". *IEEE Trans. on Signal Processing*, 39(5):1110–1121, May 1991.

[VOK91] M. Viberg, B. Ottersten, and T. Kailath. "Detection and Estimation in Sensor Arrays Using Weighted Subspace Fitting". *IEEE Trans. on Signal Processing*, 39(11):2436–2449, November 1991.

[War94] J. Ward. "Space-Time Adaptive Processing for Airborne Radar". Technical Report 1015, MIT Lincoln Laboratory, December 1994.

[War95] J. Ward. "Cramer-Rao Bounds for Target Angle and Doppler Estimation with Space-Time Adaptive Processing Radar". In *29th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, 1995.

[Wax85] M. Wax. *Detection and Estimation of Superimposed Signals.* PhD thesis, Stanford Univ., Stanford, CA, March 1985.

[Wax91] M. Wax. "Detection and Localization of Multiple Sources via the Stochastic Signals Model". *IEEE Trans. on SP*, 39(11):2450–2456, Nov. 1991.

[WJ65] J. M. Wozencraft and I. M. Jacobs. *Principles of Communication Engineering.* Wiley, New York, 1965.

[WK85] M. Wax and T. Kailath. "Detection of Signals by Information Theoretic Criteria". *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 33(2):387–392, April 1985.

[Woo53] P. M. Woodward. *Probability and Information Theory, with Applications to Radar*. Pergamon Press Ltd., London, 1953.

[WT00] M. Wax and R. Tweg. "Direction of Arrival Tracking Below the Ambiguity Threshold". *IEEE Trans. on Aerospace and Electronic Systems*, 36(2):354–363, April 2000.

[WW85] A. J. Weiss and E. Weinstein. "A Lower Bound on the Mean-Square Error in Random Parameter Estimation". *IEEE Trans. on Information Theory*, 31(5):680–682, Sep. 1985.

[WZ89] M. Wax and I. Ziskind. "Detection of the Number of Coherent Signals by the MDL Principle". *IEEE Trans. on ASSP*, ASSP-37(8):1190–1196, Aug. 1989.

[WZ98] K. T. Wong and M. D. Zoltowski. "Direction-Finding with Sparse Rectangular Dual-Size Spatial Invariance Array". *IEEE Trans. on Aerospace and Electronic Systems*, 34(4):1320–1335, Oct. 1998.

[Xu01] W. Xu. *Performance Bounds on Matched-Field Methods for Source Localization and Estimation of Ocean Environmental Parameters*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2001.

[Xu02] W. Xu. "Quantitative Ambiguity Analysis for Matched-Field Source Localization". In *Proc. 36th Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, USA, 2002.

[Zat03] M. Zatman. "Space-Time Adaptive Processing Using Sparse Arrays". In *Proc. 11th Annual Workshop on Adaptive Sensor Array Processing*, MIT Lincoln Lab., Lexington, MA, USA, 2003.

[ZF96] A. Zeira and B. Friedlander. "Direction of Arrival Estimation Using Parametric Signal Models". *IEEE Trans. on Signal Processing*, 44(2):339–350, February 1996.

[ZLS97] D. Zheng, J. Li, and P. Stoica. "One-Dimensional MODE Algorithm for Two-Dimensional Frequency Estimation". In *Proc. ICASSP 97*, 1997.

[ZS98] M. Zatman and S. T. Smith. "Resolution and Ambiguity Bounds for Interferometric-Like Systems". In *Proc.* 32^{nd} *Asilomar Conf. on Signals, Systems, and Computers*, 1998.

[ZW88] I. Ziskind and M. Wax. "Maximum Likelihood Localization of Multiple Sources by Alternating Projection". *IEEE Trans. on ASSP*, ASSP-36:1553–1560, Oct. 1988.

[ZW00] M. D. Zoltowski and K. T. Wong. "Closed-Form Eigenstructure-Based Direction Finding Using Arbitrary but Identical Subarrays on a Sparse Uniform Cartesian Array Grid". *IEEE Trans. on Signal Processing*, 48(8):2205–2210, Aug. 2000.

[ZZ69] J. Ziv and M. Zakai. "Some Lower Bounds on Signal Parameter Estimation". *IEEE Trans. on Information Theory*, 15(3):386–391, May 1969.